CHARACTERIZATION OF GENE FLOW AND GENETIC DIVERSITY OF INTERSPECIFIC TEA (*Camellia sinensis* (L.) O. Kuntze) HYBRIDS USING SIMPLE SEQUENCE REPEAT MARKERS

JOB AREBA MAANGI

MASTER OF SCIENCE
(Molecular Biology and Bioinformatics)

JOMO KENYATTA UNIVERSITY OF
AGRICULTURE AND TECHNOLOGY

2022

Characterization of gene flow and genetic diversity of interspecific tea (*Camellia sinensis* (L.) O. Kuntze) hybrids using Simple Sequence Repeat markers

Job Areba Maangi

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Masters of Science in Molecular Biology and Bioinformatics of the Jomo Kenyatta University of Agriculture and Technology

2022

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature…………………………………… Date …………………………………

Job Areba Maangi

This thesis has been submitted for examination with our approval as per the requirements of the University.

Signature……………………………… Date ……………………………………

Dr. Joel L. Bargul, PhD

JKUAT, Kenya

Signature ……………………………… Date ……………………………………

Dr. Richard Chalo Muoki, PhD

TRI, Kenya

DEDICATION

I dedicate this thesis to my dear wife, Lydiah Moraa, and our children, Stacybeth, Tracy, and Salyolivia for their love, inspiration and support that enabled me to achieve the goals of this study. I would also like to dedicate it to my dear parents, Mr. & Mrs. Maangi, my siblings, and friends for their support and encouragement. I appreciate your support and pray for God’s blessings to be upon you all.

ACKNOWLEDGMENTS

I am thankful to God in whom this work was possible. I am also grateful to the Tea Research Institute's Crop Improvement and Management (CIM) program for availing the necessary research infrastructure for my study (then headed by Dr. Samson Kamunya) and the Kenya's National Research Fund for funding this research under the 2017 postgraduate cohort.

My sincere appreciation and gratitude goes to my supervisors, Dr. Joel L. Bargul and Dr. Richard Muoki for their exemplary guidance, support, patience, and encouragement throughout my research studies.

I also extend my gratitude to the laboratory technicians at the CIM molecular laboratories for technical assistance during collection and processing of field samples. Many thanks to the library staff of Tea Research Institute (TRI) for availing literature and materials for compiling my thesis.

I would also like to recognize my colleagues, Josephat Bungei, Robert Korir, Oliver Rop, and Celestine Magiri, and all the staff at the TRI for their moral support and technical assistance during the course of my research. I appreciate your friendship and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS AND ACRONYMS

°E	Degrees East
°S	Degrees South
µl	Microliter
AFLP	Amplified Fragment Length Polymorphism
bp	Base pairs
CAPS	Cleaved Amplified Polymorphism Sequences
cDNA	Complementary DNA
CIA	Chloroform Isoamyl alcohol
CIM	Crop Improvement and Management Program
CTAB	Cetyltrimethylammonium bromide
D	Discriminating power
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleoside triphosphate
EC	Epicatechin
EGCG	Epigallocatechin-3-gallate
EGC	Epigallocatechin
EST	Expressed sequence tag
EST-SSR	Expressed Sequence Tag-Simple Sequence Repeats
EtBr	Ethidium Bromide
FIS	Inbreeding coefficient
$\mathbf{F_{ST}}$	Fixation index
He	Expected Heterozygosity
Ho	Observed Heterozygosity
ISSR	Inter Simple Sequence Repeats
MAS	Marker-assisted selection
NCBI	National Center for Biotechnology Information
mM	Milimolar
NaCl-TE	Sodium chloride - Tris EDTA buffer

ng	Nanograms
NR	Non-redundant sequences
PAL	Phenylalanine Ammonia-Lyase
PCR	Polymerase Chain Reaction
PIC	Polymorphism Information Content
PolyA/T	Poly Adenine/Thiamine tail
RAPD	Random Ampliiified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphisms
RNA	Ribonucleic Acid
RNase	Ribonuclease
SNPs	Single Nucleotide Polymorphisms
SSRs	Simple Sequence Repeats
Tm	Melting temperature
TRFK	Tea Research Foundation of Kenya
UV	Ultra Violet
V	Volts

ABSTRACT

Tea [*Camellia sinensis* (L.) O. Kuntze] is an evergreen, economically important crop in Kenya and globally that is characterized by high genetic variability, from which tea, a popular soft beverage that is widely consumed, is produced. Tea improvement depends on the extent of genetic diversity within the available population and the ability of the tea crop to hybridize freely within the species as well as with closely related 'wild' species in the genus *Camellia*. There are prospects of using interspecific hybridization for introducing desirable traits in tea such as cold hardiness, drought tolerance, and specific characters in chemical components, and disease and pest resistance, among others. However, the contribution of wild species to the cultivated gene pool is presently not well understood. This study characterized genetic diversity and gene flow in interspecific tea hybrids by genotyping SSRs across eight loci and analyzing the levels of relatedness in the population. Twenty SSR markers comprising of five novel EST-SSRs and fifteen adapted microsatellites were initially screened for polymorphisms using three randomly selected interspecific hybrids (i.e. TRFK 570/2, TRFK 688/1, and TRFK 83/1) and one intraspecific cultivar (TRFK 6/8). Of these, eight most informative polymorphic microsatellites were used to study genetic diversity and gene flow in 88 tea accessions comprising interspecific hybrids, parental clones, and wild tea species. DNA was extracted from each genotype and SSR fragments were PCR-amplified, separated on 1.5% agarose gel, and binary data (1= present, 0= absent) scored at the eight loci. The polymorphic information content (*PIC*) and discriminating power (*D*) of SSR markers were determined using the iMEC program, whereas genetic diversity in the genotypes was estimated with POPGENE version 1.32. Analysis of molecular variance was performed using GenAlEx 6.5 software while parentage analysis was performed with Cervus 3.0.7. GenStat (15th edition) and Structure 2.3.4 were used to analyze relatedness based on Jaccard's coefficient and genetic structure of the population, respectively. Eight markers were relatively informative, with *PIC* and D values of 0.40 and 0.30, respectively. Genetic diversity was highest in Genet 3c/2007 population (*Ne* = 1.9727 and *I* = 0.6862) and lowest in wild tea population (*Ne* =1.4320, *I* =0.4105) that occur as isolated pure wild type groups with reduced genetic exchange with other *Camellia* species. Among the families, St. 645 was the most diverse (*I* = 0.64) and St. 31 the least diverse (*I* = 0.36). The population was only moderately differentiated (FST = 0.0661) across the eight loci, suggesting past genetic exchanges. The close relatedness among the accessions was revealed by neighbor-joining analysis with most hybrids clustering in a manner consistent with known pedigree information. Wild alleles were highest in Genet 3c/1999 hybrids (95%) and lowest in Genet 3c/2005 hybrids (38.9%) demonstrating a relatively high but unequal genetic contribution of wild *Camellia* species into cultivated tea under natural pollination conditions. Parentage analysis showed multiple and shared paternity among half-sib and full-sib families. As the results demonstrate that EST-SSRs are highly efficient in identifying interspecific tea hybrids, typing more EST-SSR loci could be useful for accurate parental reconstruction in progenies with unknown identity and half-sibs from polycross mating and for the determination of genetic diversity patterns in tea breeding stocks for hybridization breeding.

CHAPTER ONE

INTRODUCTION

1.1 Background

Tea [*Camellia sinensis* (L.) O. Kuntze] is a popular soft beverage that is consumed globally as green, black, yellow, Oolong, or white tea, which are distinguished based on the aeration level during processing (Fang *et al*., 2014). The genus *Camellia* belongs to Theaceae family that is indigenous to Central Asia and has over 320 reported species that naturally hybridize (Mondal, 2011; Mukhopadhyay *et al*., 2016). From its Central Asia origins, tea is currently cultivated in diverse environments, ranging from 49ºN to 30ºS and altitudes from sea level to 2700m (Zhen *et al*., 2005). Tea is an economically important crop, as it is a leading foreign exchange earner for countries in Asia and Africa, including Kenya, where it contributed USD 1.098 billion to the Kenyan economy in 2019 (ITC, 2019). Tea sales account for approximately 26% of export earnings and contribute about 4% to Kenya's GDP annually (Muoki *et al*., 2020). Further, more than 750,000 farmers directly earn a living from tea and over 6 million Kenyans directly or indirectly depend on it for their livelihoods (Muoki *et al*., 2020).

Historically, tea cultivars are progenies of diverse seed sources that were subsequently vegetatively propagated (Chen *et al*., 2005). However, a recent breeding strategy involves artificial pollination and hybridization among selected tea accessions or with wild *Camellia* relatives, resulting in diverse intraspecific and interspecific hybrids (Wachira *et al*., 1997; Kerio *et al*., 2012; Wachira *et al*., 2013). Wild species have been historically used as sources of genetic variation in crop improvement. As such, gene flow involving wild species and their domesticated counterparts is valuable for the enrichment of the effective breeding population. Dispersal of tea led to varietal speciation and the evolution of three distinct cultivated taxa namely: var. *sinensis* ('China tea'), var. *assamica* (Masters) Kitamura ('Assam tea') and var. *assamica* spp. *lasiocalyx* (Planchon ex Watt) ('Cambod tea') (Preedy, 2012). These three cultivated taxa are differentiated based on their morphological (including foliar, floral, and growth features), biochemical, and molecular characters (Wachira *et al.*, 2013). However, the occurrence of pure commercial archetypes of tea is unlikely because of overlapping characteristics produced from extensive hybridization among the three commonly cultivated varieties (Assam, China, and Cambod teas) and interspecific crosses with other *Camellia* species (Banerjee, 1992; Ming & Bartholomew, 2007; Kamunya *et al.*, 2012). The knowledge of gene flow and genetic variability in local interspecific hybrids could be useful in the identification of superior genotypes for the enrichment of effective tea breeding populations (Banerjee, 1992). Therefore, the selection and crossing of cultivated tea with wild populations can be used to generate potentially high yielding interspecific varieties such as purple tea cultivars (Chahal & Gosal, 2002). Wild tea species have been shown to improve some key traits, for instance, cold hardiness, drought tolerance, specific characters in chemical components, disease and pest resistance (Preedy, 2012). For such purposes, extensive collections of tea germplasm have been made at the Tea Research Institute, Kenya (Wambulwa *et al.*, 2016).

Molecular markers are useful tools in crop improvement that are utilized in the speedy development of superior varieties through marker-assisted selection (MAS). DNA-based markers have also been applied to identify tea varieties from a broad range of commercial tea products (Stoeckle *et al.*, 2011). In particular, they have been used to differentiate morphologically indistinguishable tea varieties (Freeman *et al.*, 2004; Yao *et al.*, 2005; Wambulwa *et al.*, 2016). The markers are also efficient in studying genetic relationships as they are reproducible, multi-allelic, informative, polymorphic, relatively abundant in the genome, and co-dominantly inherited (Navajas & Fenton, 2000; Gupta *et al.*, 2005). The number of microsatellites or Simple Sequence Repeats (SSRs) in the genome changes rapidly during evolution and being co-dominant markers; they can distinguish homozygote from heterozygote genotypes (Oliveira *et al.*, 2006). Additionally, SSR fingerprints are useful in evaluating the gene flow in hybridization events that produce interspecific hybrids for cultivation (Mondal, 2002). Other molecular marker such as sequence-tagged sites (Wachira *et al.*, 2001) and cleaved amplified polymorphic sequences (CAPS) have also been used to discriminate tea varieties (Kaundun & Matsumoto, 2003). Several improved interspecific cultivars such as TRFK 306/1, a colored interspecific hybrid of *C. irrawandiensis* (wild tea), have been released for commercial cultivation (Wambulwa *et al*., 2016). As the level of gene introgression between the cultivated tea and its wild relatives has never been examined, the present study investigated the gene flow and genetic diversity in interspecific populations of tea using Expressed Sequence Tags-Simple Sequence Repeat (EST-SSRs) markers.

1.2 Statement of the Problem

Due to late-acting self-incompatibility where self-pollen tubes fail to penetrate and fertilize ovules, breeding pure lines in tea is not possible (Wachira & Kamunya, 2005; Chen *et al.*, 2012). However, the prospect of using the wild genetic stocks in introducing some desirable traits, such as cold hardiness, drought tolerance, specific characters in chemical components, disease and pest resistance, among others, exists. Over-exploitation of fewer genetically outstanding breeding stocks in breeding programs has narrowed the genetic base of cultivated tea, reducing the fitness and performance of commercial cultivars in different agro-ecological zones. In Kenya, selective breeding has involved crossing a few elite parents, which has resulted in the narrowing of the genetic base. Notably, of the 45 clones released for commercial cultivation by the Tea Research Institute, 60% of them are progenies of clone 6/8 (Kamunya *et al*., 2012). In addition, only three clones with high-yielding potential, namely, clones 6/8, S15/10, and BB35 and two with lower susceptibility to drought, namely, 31/8 and TN 14-3, are mostly commercially cultivated in the country. Thus, there is a risk that the genetic bases of the breeding stocks and commercial clones are narrowing. Wild *Camellia* species have historically been useful sources of genetic variation for tea improvement programs. Because of the close morphological resemblance between many *Camellia* species, it is possible that several wild *Camellia* species and their hybrids with tea have remained undetected in tea fields. In an effort to access diversity from the secondary and tertiary gene pools of tea, several *Camellia* species were imported into Kenya and conserved in open fields. These include

C. japonica, C. brevistyla, C. sasanqua, C. irrawadiensis, C. assimilis, C. oleifera, C. kissi, C. chrysantha, C. furfuraceae, among others, that were planted out in a '*Camellia* Gene Bank'. However, the extent of genetic contribution of these taxa to the cultivated tea germplasm in Kenya is little understood, despite 105 putative hybrid progenies being developed in three separate experimental trials in 1999, 2005 and 2007. The present study investigated the gene flow and genetic diversity in interspecific hybrid populations of tea.

1.3 Justification

Molecular markers such as microsatellites and Single Nucleotide Polymorphisms (SNPs) are being rapidly adopted for crop improvement as an effective and appropriate tool for assessment of genetic diversity and trait-specific crop improvement (Bandyopadhyay, 2011). Simple Sequence Repeat (SSR) markers have been used for discriminating and assessing the genetic purity of parental lines and hybrids in crops like rice (Yashitola *et al*., 2002; Nandakumar *et al*., 2004), maize (Mingsheng *et al*., 2006), and sunflower (Antonova *et al*., 2006). Unlike other markers, SSRs are simple, highly polymorphic, multi-allelic, and co-dominantly inherited and existing abundantly in both intronic and exonic genomic regions (Gupta *et al*., 2005).

Interspecific hybridization represents an important process towards product diversification and evolutionary studies in tea (Bandyopadhyay, 2011). For example, anthocyanins were recently introduced in cultivated varieties through interspecific hybridization (Gasura *et al*., 2008). Worldwide, the tea plant has received immense attention due to its proven pharmacological properties. With a long history of development and cultivation of interspecific hybrids, Kenya is home to broad secondary and tertiary gene pools of *Camellia* species (Kilel *et al*., 2013). Although these accessions are used in tea improvement, their genetic contribution towards cultivated species has not been quantified (TRFK, 2012). An understanding of the contribution of wild *Camellia* species, would provide an informative scientific basis for broadening the current germplasm collections for breeding and conservation activities (Wachira *et al*., 1995). It would also help in the identification of wild parental lines for inclusion in tea breeding so as to maintain a wide tea genetic base to mitigate against climate related challenges (TRFK, 2012). The present study analyzed polymorphic SSR markers in *in-situ* intra- and inter-specific tea collections at the TRFK 'Gene Bank' to determine the genetic diversity and contribution of wild tea species to the cultivated tea gene pool.

1.4 Null Hypothesis

There is no difference in gene flow and genetic diversity between interspecific tea hybrids and wild *Camellia* species.

1.5 Objectives

1.5.1 General Objective

To characterize the gene flow and genetic diversity in interspecific tea hybrids using SSR markers.

1.5.2 Specific Objectives

1. To evaluate the use of EST-SSR markers and genomic SSRs in identification of interspecific hybrids of tea.
2. To characterize the genetic diversity of interspecific hybrids from selected Tea Research Foundation of Kenya tea germplasms using SSR markers.
3. To determine the relative genetic contribution of wild tea species to the gene pool of cultivated teas using SSR markers.

1.6 Research Questions

1. How useful are novel EST-SSR markers in discriminating interspecific hybrids of tea compared to genomic SSR markers?

2. What is the genetic diversity of interspecific hybrids from selected Tea Research Foundation of Kenya tea germplasms?

3. What is the relative genetic contribution of wild tea species to the gene pool of cultivated teas?

CHAPTER TWO

LITERATURE REVIEW

2.1 Botany of Tea

Tea [*C. sinensis* (L.) O. Kuntze] is a non-alcoholic, caffeine-rich beverage widely consumed for its attractive aroma, medicinal value, and mildly stimulating effects (Karak & Bhaghat, 2010). Free-growing tea trees can reach 20-30 meters when unpruned and survive for about 100 years, but for cultivation purposes they are maintained at a height of 1-2 meters (Kamunya *et al*., 2019). This perennial crop is kept evergreen by pruning at an interval of 2-6 years, depending on the climate of the tea growing region (Willson & Clifford, 1992). Unpruned trees have fewer leafy flushes annually. After pruning, shoots that develop from leaf axils are plucked every 7-14 days until the growing season ends (Barua, 1970). Plucked shoots comprise 2-3 leaves and an apical bud that are utilized to process tea (Kamunya et al., 2019).

Botanical classification places tea in the genus *Camellia*, which has over 200 species (Wachira *et al*., 2013). Early classification by Sealy (1958) comprised 12 sub-generic groups, including Thea under which cultivated tea belonged. Later 24 other previously unknown species were discovered, which led to a revision of Sealy's classification. Four subgenera were now recognized under the genus *Camellia*, namely, *Protocamellia*, *Camellia, Thea*, and *Metacamellia*, along with 20 sections (Chang & Bartholomew, 1984).

Linnaeus (1753) first named tea scientifically as *Thea sinensis*. The Linnaeus classification was revised after two morphologically distinct groups of tea were identified in Assam-Tibet region, namely, *Thea sinensis* (small-leaved) and *Thea assamica* (large-leaved) (Masters, 1844). *Thea* and *Camellia* remained separate taxa until the mid-1900s when some researchers considered the morphological and biochemical differences as natural variation in leaf characters (Wachira *et al*., 2013). *Thea* was considered synonymous to *Camellia* but *Camellia* was agreed upon as the generic name (Wright, 1962). Today, tea is botanically called *Camellia sinensis* (L.). O. Kuntze, regardless of intraspecific differences (Wachira *et al.*, 2013). An ideal habitat for tea plant is shaded areas, an altitude of 2100-2700m, tropical and subtropical climates receiving 1200-2200mm of rainfall that is distributed throughout the year, temperatures of 13°C-30°C (Kamunya *et al.*, 2019). Tea also requires windbreaks that lower evapotranspiration; hence, tea plantations are located at the edge of forests or belts of tall tree species such as *Hakea saligna* and *Grevillea robusta* (Willson & Clifford, 1992). The tea plant requires deep well-drained red volcanic soils that are slightly acidic (pH = 4-5.6) (Kamunya *et al.*, 2019). It is a perennial crop with diverse morphological traits, genetics, and a long history of cultivation and distribution.

2.2.1 Morphology of Tea

Tea is a leafy, perennial, out-crossing plant that can naturally grow to a height of up to 30m though maintained at 0.6-1.5m under cultivation (Yamamoto *et al*., 1997). It produces solitary or paired white fragrant flowers at the axils and four-seeded green fruits after 1–6 years (Yamamoto *et al*., 1997). The fruits are brownish green and encase 1-4 spherical seeds (Mahmood et al., 2010). The flowers are scented and appear singly or in clusters of 2-4 on short stalks in the leaf axils (Kamunya *et al.*, 2019). Each flower is about 4cm in diameter with five sepals and 5-9 petals and is hermaphroditic (contains both male and female parts) (Barua, 1970). The stamens are many and organized in whorls, with shorter inner ones and elongated outer ones about 9-13mm in length and joined at the base with sepals (Syahbudin et al., 2019). The flower contains a free style that are usually three and a hairy ovary bearing 3-5 ovules (Ross, 2005).

Leaf morphology is the main criterion for distinguishing the major tea taxa. Three major tea varieties have been distinguished based on leaf morphology: small-leaved China tea (*C. sinensis* var. *sinensis*), large-leaved Assam tea (*C. sinensis* var. *assamica*), and Cambod tea (*C. sinensis* var. *assamica* spp. *lasiocalyx*) (Figure 2.1) (Barchetia *et al*., 2009). The leaves are usually light green (young leaves) or bright green (mature leaves), coriaceous, lanceolate with serrated margins and are 5 – 30 cm long (Mahmood *et al*., 2010). For Assam tea, leaf blade is broad, less erect, and elliptic in shape, 8-20 cm in length and 4-8cm in width, with few serrations, and is less erect (Wachira et al., 2013). In contrast, China tea have small erect leaf blade and serrated leaf margins with a broadly obtuse apex and a petiole that is stout, 3-7mm long, giving the leaf an erect position (Barchetia *et al.,* 2009). Cambod tea has an intermediate leaf size between China tea and Assam tea. The leaves are broad and elliptic, more or less erect and light green (Syahbudin et al., 2019). The fruit is compact, smooth, and rounded three-compartmentalized capsule, bearing solitary seeds in each compartment (Biswas, 2006).

The flush shoot (apical bud and 2-3 leaves) is picked weekly or fortnightly, depending on the variety and climatic conditions (Yamamoto *et al*., 1997). During processing, varying fermentation levels produces different teas, such as green, black, and Oolong teas. However, these taxonomic groups can freely interbreed leading to high genetic heterogeneity (Heiss & Heiss, 2007). Vegetative propagation is used to upscale superior hybrid seed progenies forming clonal teas that are further tested and later released for commercial cultivation (Korir *et al*., 2013). These clones are morphologically distinguishable based on foliar, leaf, and fruit shapes (Lai *et al*., 2001).

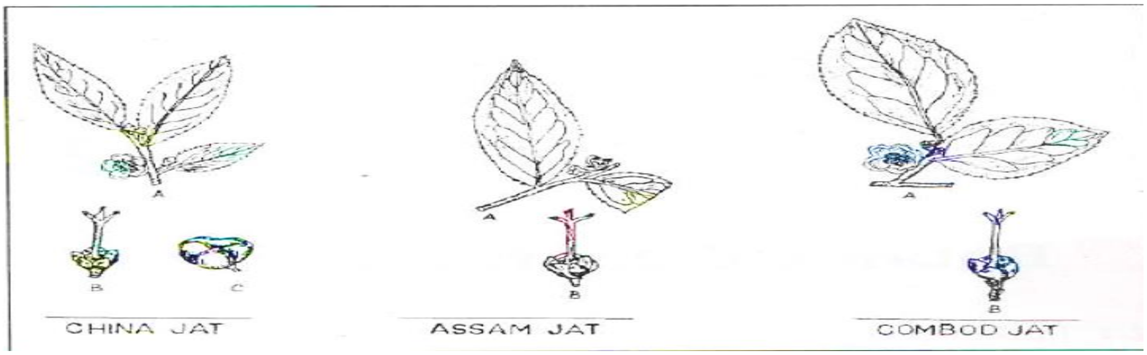


Figure 2.1: Leaf morphology of three distinct groups of tea (Barchetia *et al.*, 2009).

2.2.2 Genetics of Tea

Cytogenetic studies have revealed the chromosomal biology of the tea plant. Xia *et al.* (2020) reported that karyotyping of tea found 15 chromosomes in *C. sinensis* gametes, which suggested that 30 chromosomes occur in diploid tea plants (2n = 30). Similar findings were reported in various clonal tea accessions obtained from different parts of the world (Sheidai et al., 2004; Furukawa et al., 2017). Other species in the genus *Camellia*, such as *C. oleifera* and *C. sasanqua*, exhibit polyploidy with chromosome numbers varying from 45 to 120 (Huang *et al.,* 2013).

Tea has a relatively large genome size that is estimated to be 3.8-4.0 Gb (Hanson *et al.,* 2001; Tanaka *et al.,* 2006). However, the genome size of two varieties, shuchazao and yukang#10, was estimated at 3.0 Gb upon sequencing (Xia et al., 2017). Therefore, the genome size appears to be largely conserved but variations occur between tea varieties and *Camellia* species because of past hybridization events (Huang *et al.,* 2013).

Tea is characterized by self-incompatibility, a trait attributed to the high genetic heterogeneity in the crop (Wachira & Kamunya, 2005). Pollen tubes in self pollens extend through the style and enter the ovary but fail to penetrate the ovules, a phenomenon called late-acting self-incompatibility (Chen *et al.,* 2012). This trait favors cross-pollinations, resulting to highly heterogeneous intraspecific and interspecific hybrids.

Sequencing projects have revealed specific genes encoding secondary metabolites associated with quality tea characteristics such as aroma or taste, as well as resistance to drought and pests and diseases (Dodds & Rathjen, 2010). In total, 33,932 protein-coding genes occur in *C. sinensis* var. *sinensis* (China tea) and 36,951 occur in *C. sinensis* var. *assamica* (Assam tea) (Xia *et al.,* 2020). Most of these genes encode enzymes involved in biosynthesis of secondary metabolites. For example, serine carboxypeptidase-like acylltransferase plays a role in the synthesis of galloylated catechins that is an important marker of tea quality and taste (Wei *et al.,* 2018). Genetic drift could account for the differences in number of functional genes in the two genomes. The two tea varieties are thought to have diverged from a common progenitor about 0.38 to 1.54 million years ago (Xia *et al.,* 2017). The adaptability of tea to diverse agroecological zones globally is attributed to duplicated disease resistance genes and pattern-recognition receptors (Xia *et al.*, 2020).

2.1.3 Origin and Distribution of Tea

Tea has a long history of cultivation from its wild progenitors and use as a beverage. The earliest textual evidence of consumption of wild teas can be found in China, where it was exploited as medicine during the Shang Dynasty (2737 BC) and later as a beverage during Zhou Dynasty (1000 BC) (Chang & Bartholomew, 1984). Tea was first domesticated over 3,000 years ago in Chinese regions (Yamanishi, 1995). This is corroborated by archaeological evidence showing that tea consumption was common among emperors in Han dynasty that existed over 2,000 years ago (Lu *et al.*, 2016). In addition, the origin of the tea plant is believed to be Yunnan province, China, the native habitat of the Pureh tea variety, where ancient trees as old as 1700 years still grow (Chang & Bartholomew, 1984). This evidence shows that the China tea variety was first domesticated in China but the original birthplace remains unknown. In southern China, wild tea species grow naturally as perennial forests in various areas of Yunnan and Gandong provinces (Chang & Bartholomew, 1984). The archaeological evidence and occurrence of wild tea varieties strongly suggest that tea is native to China.

Outside China, many regions are plausible historical centers of the tea domestication, including the Indian region of Assam, and the Indo-Burmese border region (Lu, 1974). This is supported by the discovery of wild tea plants indigenous to Assam, northeastern India, and Burma (Chang & Bartholomew, 1984). Therefore, these regions are a part of the original centers of the tea, which is classified as Assam tea. A recent investigation using nuclear microsatellites showed two distinct domestication origins of tea: China and India, which is consistent with the two main tea taxa, China tea and Assam tea (Meegahakumbura *et al.*, 2016). The origins of tea can be considered to be a fan-shaped Central Asia region encompassing areas in India, Burma, China, Thailand, and Vietnam between 95°-120°E and 11°-29°N (Harler, 1964).

From its original domestication centers in Indo-Chinese region, tea cultivation first spread to Indonesia, where trees were grown the island of Java in the late 1600s (Wight, 1959). Commercial plantations were first established in Japan in early 1800s using China tea seedlings and later Assam types were introduced in 1878 (Lu *et al.*, 2016). Tea was introduced in Sri Lanka in the 1860s as a substitute to coffee that was highly susceptible to disease (Lu *et al.*, 2016). In Africa, tea was first planted in Malawi in 1885, with the first plantation being established six years later. In East Africa, the cultivation of tea was started in the 1900s in Kenya, Uganda, and Tanzania. Tea was first introduced in Kenya by G. Caine in 1903 and planted on experimental basis in the present-day Limuru region, Kiambu County (TRFK, 2012).

Presently, tea is cultivated in more than 50 tropical and subtropical countries of Asia, Africa, and South America (Wambulwa *et al.*, 2017; Karunarathna *et al.*, 2018). The tea growing zones are diverse environments, ranging from 49ºN to 30ºS and altitudes from sea level to 2700m (Zhen *et al.*, 2005).

2.1.4 Economic Importance of Tea

Globally, tea is an economically important crop that is a major foreign exchange earner for producing countries. It is the most widely consumed nonalcoholic beverage after water, and its popularity is projected to grow by 2% yearly, across Asia, European Union, and Arab countries (Hicks, 2001; FAO, 2019). The increase in consumption is linked to income growth in the main markets and the production of teas healthier than coffee or cocoa (Dutta, 2017). In 2017, total tea exports globally were 1.91 million tons, after consistent growth of 2.1% over the past decade (FAO, 2019). Presently, the largest exporter of tea is China, with tea export earnings of about $2.04 billion in 2020, followed by Sri Lanka ($1.33 billion) and Kenya ($1.22 billion) (Rider, 2022).

To meet the high global demand, tea production by leading producers has grown exponentially over the past few decades. Globally, the tonnage of tea produced rose 2.5 times to 6.34 million tons in 2019 from the 1990 level (FAO, 2019). Among all tea types, the demand for black tea is the highest globally at 1.4 million tons, equivalent to 78% of all tea exports in 2017 (International Tea Commitee, 2018). Kenya is the leading exporter of black tea and its exports have almost doubled over the past three decades (Xu *et al.*, 2022).

Both the acreage under tea in Kenya and the country's earnings from exports have increased over the past few decades. In 2018, total tea production in Kenya was 493 million, which earned the economy Kshs. 140 billion (about $1.30 billion) (Muoki *et al.*, 2020). This amount accounted for 26% of the country's total earnings from all exports and an equivalent of 4% of the GDP (International Tea Committee, 2018). The leading markets for Kenyan tea include Pakistan, Egypt, the UK, and Sudan, cumulatively accounting for 62% of exports (Tea Board of Kenya, 2010).

In addition to being a leading foreign exchange earner for Kenya, the tea industry supports millions of livelihoods, especially in rural areas where tea is mainly produced. Tea production in Kenya is largely rural based, where 62% of all tea is produced by small-scale farmers, directly supporting about three million people (Tea Board of Kenya, 2010). The total acreage under tea cultivation in Kenya is estimated to be 232,742 ha in 18 counties, including Kericho, Nyeri, Kiambu, Nandi, and Kisii, and with reduced mechanization, 10% of the population earn its livelihood directly or indirectly from tea (International Tea Committee, 2018).

The contribution of tea to the rural economy and reduction of rural-urban migration is therefore significant (Wachira, 2002). The tea industry has contributed to infrastructural development in rural areas, including roads and schools, and supported environmental conservation efforts through decreased soil erosion in tea plantations and mitigation of climate change (Muoki *et al.,* 2020). Therefore, sustainability the tea industry is important to the economic growth and development of Kenya both as foreign exchange earner and source of livelihood for millions of people that depend on the crop directly or indirectly in tea growing zones across the country.

2.1.5 Challenges of Tea Production

The global tea industry experiences many challenges related to production, markets, and resource constraints. In tea production, the main constraint include high costs since tea production is labor intensive (Mwangi, 2016). The workforce requirements during ploughing, land preparation, nursery development, planting, and maintenance, including weeding, mulching and pruning to maintain a height of about one meter, are high (Onduru *et al.,* 2012). Non-mechanized plucking by hand is labor intensive. Other highly labor demanding processes during non-mechanized tea processing steps, including steaming, drying, grading, and packaging (Mwangi, 2016). Major tea companies have a huge labor force earning a daily wage of about $1.5 but smallholder farms depend on unpaid family members (Onduru *et al.,* 2012). The high cost of labor impacts sustainable production of tea and profitability of tea companies. Additional costs come from transportation of plucked tea to factories, fertilizers, and taxes that affect earnings, especially for the majority small-scale tea growers (Gesimba *et al.,* 2005). In India, the estimated cost of production is $2,170 per hectare for smallholder farmers, and it includes the cost of procuring cuttings, hiring labor, irrigation, and purchasing weedicides, insecticides, and mulch (Das & Mishra, 2020).

The persistently low export prices is also a challenge to the sustainability of the tea industry. Consistent expansion of acreage under tea has increased global tea export, which creates a glut and pushes the prices down (Gesimba *et al.,* 2005). The price of the commodity is also externally determined, with multinational companies (MNCs) and private firms such as Finlays and Unilever that dominate the tea industry manipulate tea supply and pricing (Ndege, 2021). These companies dictate the tea types, quantity, quality, and prices of teas entering the international market. MNCs control tea auctions where 70% of tea is traded globally; thus, they can manipulate prices through intermediaries and determine the earnings of small-scale tea growers (Ndege, 2021). The stagnant prices have seen India establish a price stabilization fund to protect smallholder farmers (Das &

Mishra, 2020). In Kenya, tea prices fell by 12% between 2020 and 2021, partly due to a global glut in tea supply (Ndege, 2021).

Pests and diseases pose a major challenge to tea production. Several pests and diseases attack foliage, stems, and roots, which affects the growth, yield, and quality of tea (Pandey *et al.,* 2021). Fungal diseases such as blister blight are the most prevalent in tea plantations and have contributed to significant economic losses of 20-50% in India and Indonesia (Gulati *et al.,* 1993; Radhakrishnan & Baby, 2004). Fungal infections also lower the quality of tea by reducing the levels of caffeine, catechins, and aromatic compounds (Murr *et al.,* 2015). Other diseases that contribute to a decline in yield and quality include anthracnose, gray blight, stem cankers, and root rots (Pandey *et al.,* 2021). In Kenya, Armillaria root rot was associated with 50% loss in output in small-scale farms (Onsando *et al.,* 1997). Increasing temperature due to climate change may increase losses due to diseases and pests (Muoki *et al.,* 2020).

Frequent droughts due to global warming presents a serious challenge to tea production globally. As water resources decline, the focus has turned on breeding drought tolerant varieties (Muoki *et al.*, 2020). Advances in breeding have led to the release of new high-yielding clonal tea, forcing farmers to uproot old well adapted seedling tea plantations, which presents a challenge to maintaining on-farm diversity (Kamunya et al., 2012). Another problem is the overreliance on few breeding stocks, such as clone TRFK 6/8, which accounts for 67% of all teas grown in Kenya (Wachira, 2002).

2.2 Genetic Diversity of Tea

The genus *Camellia* encompasses over 325 species up from 200 species in the 1980s (Mondal, 2011; Mukhopadhyay *et al*., 2016). Currently, there are about 2500 cultivated varieties worldwide with diverse traits such as disease resistance (blister blight), water stress/frost tolerance, and caffeine content as well as leaf color, pose, and pubescence (Mondal, 2011; Bramel & Chen, 2019). The Tea Research Institute, formerly the Tea Research Foundation of Kenya (TRFK), has developed over 58 tea varieties for cultivation

(TRI, 2019). In general, three main taxa contribute to the gene pool of tea, namely, *C. sinensis* var. *assamica*, *C. sinensis* var. *sinensis*, and *C. sinensis* var. *assamica* sub sp. *Lasiocalyx* (Barchetia *et al*., 2009). However, a high degree of introgression between tea species yields numerous hybrids with a broad continuum of morphological traits between the Assam and Chinese archetypes (Barchetia *et al*., 2009). These cultivated teas naturally hybrid with their wild relatives, resulting in highly heterogeneous interspecific hybrids (Heiss & Heiss, 2007).

2.3 Breeding and Selection of Tea

Tea is naturally cross-pollinated. Field selection for superior traits is a common practice in commercial tea farming (Mondal, 2002). Elite plants developed from existing archetypes, namely, *C. sinensis* var. *assamica, C. sinensis* var. *sinensis*, and *C. sinensis* var. *assamica* sub. sp. *Lasiocalyx*, are selected and multiplied through vegetative propagated (Heiss & Heiss, 2007). However, since selection is based on optimum yield, quality, and resistance to biotic and abiotic stresses, genetic erosion is likely to occur unless clones of disparate origin are used (Mondal, 2002). Plantation cultivation of clonal tea further reduces the genetic diversity over time (Bandyopadhyay, 2011). Although conventional breeding by crossing selected tea types has led to tea improvement, genetic bottlenecks, such as inbreeding depression, vulnerability to stress, long gestation periods, long seed maturation period, and variation in flowering time between clones, hamper the prospect of improving desirable traits in tea (Mondal *et al*., 2004). Therefore, seed-grown tea plants, which display a high heterogeneity, are the viable options for developing improved tea varieties prior to multiplication through vegetative propagation and grafting (Bandyopadhyay, 2011).

2.4 Interspecific Tea Hybrids

Pioneer tea plantations were established from heterogeneous seeds obtained from India and China, whereas later plantations were established from clonal teas selected for high yield and quality (Kamunya *et al*., 2010). Interspecific hybrids are either half-sib (open pollinated) or full-sib (controlled cross-pollinated) progenies of crosses between *C. sinensis* with related wild tea species, such as *C. japonica, C. taliensis,* and *C. irrawandiensis*. However, several stocks of interspecific hybrids established at the TRI with unknown paternity (TRFK, 2012). Some of these interspecific hybrids have purple-colored leaves attributed to rich anthocyanin content that make them suitable for tea products diversification (Kamunya *et al*., 2012; Kilel *et al*., 2013). Moreover, biochemical analyses show that purple tea products are richer in polyphenols (Karori *et al*., 2007) and catechins such as epicatechin (EC), epigallocatechin-3-gallate (EGCG), epigallocatechin (EGC) (Lai *et al*., 2016) than green leaf tea cultivars. Further, the hybrids have also been shown to contain lower caffeine content compared to green tea cultivars (Kilel *et al*., 2013). Examples of the colored hybrids cultivated in Kenya include; TRFK 306, TRFK 73/1, TRFK 73/2, TRFK 73/3, TRFK 73/4, TRFK 73/5, TRFK 73/7, and TRFK 83/1 (Kilel *et al*., 2013). Though these cultivars are classified under one taxon, based on their pigmented leaves, the genetic variability and wild-to-tea gene flow has not been studied.

2.5 Assessment of Genetic Diversity in Tea

Genetic diversity refers to the genetic variation within a taxon, i.e., population, genus, or species (Chen *et al*., 2005). There is varying tea diversity in different growing regions owing to the cultivation of genetically diverse varieties (Olson *et al*., 1995). However, selection for specific traits of interest may narrow the genetic variability among cultivated varieties compared to their wild progenitors (Olson *et al*., 1995). Thus, continued development of high-yielding varieties poses a threat to tea genetic diversity (Khlestkina *et al*., 2004).

Tea diversity has been assessed using morphological descriptors (Chen *et al*., 2005), biochemical components (Magoma *et al*., 2000; Chen *et al*., 2005), and allozymes (Yeeh *et al*., 1996; Chen *et al*., 2005). Other studies have used molecular markers such as Restriction Fragment Length Polymorphisms (RFLPs) (Matsumoto *et al*., 2002; Devarumath *et al*., 2002), Random Amplified Polymorphic DNA (RAPD) (Wachira *et al*., 1995), Amplified Fragment Length Polymorphisms (AFLPs) (Balasaravanan *et al*., 2003), and microsatellites or simple sequence repeats (SSRs) (Mondal, 2002) to distinguish closely related germplasms.

2.6 Morphological Markers

Morphological markers include a set of descriptors for a species (Benjamin et al., 2008). Using Principal Component Analysis, a specific number of key phenotypic descriptors can be identified that explain the variation observed in crops in a rapid and efficient manner (Bekele & Bekele, 2014). A descriptor is a feature or phenotypic trait of a species that is quantifiable (Heywood, 1967). Morphological descriptors are not exactly equivalent for comparison purposes but offer key advantages, such as ease of observation, availability and practical application in the identification and classification of organisms (Bekele & Bekele, 2014). Some constant characters are quite stable and remain unchanged by the environment and are heritable. These are useful and cost-effective tools for identifying and cultivars and diversity studies compared to molecular markers. Homologous structures that evolved through similar pathways, somatic structures such as roots and leaves as well reproductive structures, and patterns of plant development are all useful in morphological characterization (Donald, 2001).

Morphological markers have been used by plant breeders to characterize tea and develop superior cultivars. The parts of a tea plant used in morphological characterization for breeding purposes include the leaf, stem, and branches (Magoma *et al.,* 2000). The outbreeding nature of tea results in high heterogeneity and continuous variation of morphological characters (Preedy, 2012). Leaf color, shape, size, and leaf area index have been applied in tea classification (Wachira *et al.,* 1995; Kaudum & Matsumoto, 2002; Magoma *et al.,* 2000). Leaf thickness and length and hairy buds were used to classify seven elite clones grown in Lawu mountain slopes, Indonesia (Syahbudin *et al.,* 2019). Thuvaraki *et al.* (2017) used characterized hybrid progenies based on five morphological characters: petiole pigmentation, leaf shape, pubescence, leaf color and petiole coloration. The hybrids exhibited significant variation, with 40 individuals clustering with the maternal parent (TRI 2043) and 21 individuals grouping with the paternal parent (TRI

3055). Several other morphological characters can be used for tea taxonomy and diversity analysis. Leaf size, leaf length-width ratio, internode size, bud size, petiole size, serrations at leaf margins, and shoot density are useful and stable descriptors of tea varieties (Bekele & Bekele, 2014).

2.7 Molecular Markers

Molecular markers are detectable heterozygous sites or loci based on the amino acid or nucleotide polymorphisms that can be used to distinguish closely related genotypes (Fang *et al*., 2014). Molecular markers, unlike morphological or biochemical markers, are less prone to environmental influences and can detect polymorphisms at an early stage of plant growth and development (Prince & Parks, 2001). As a result, they have gained useful applications in the fields of phylogeny, taxonomy, evolutionary studies, and breeding. Other specific characteristics that make them more preferred to morphological and biochemical markers include high polymorphic information content, co-dominant inheritance (can distinguish homozygous from heterozygous traits), abundant distribution in the genome, high reproducibility, and loci specificity (Varshney *et al*., 2005; Weising *et al*., 2005). Although molecular markers such as Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), inter-simple sequence repeat polymorphism (ISSR), SSR, and single nucleotide polymorphism (SNP) can be screened from high quality genomic DNA (Matsumoto *et al*., 1994; Wachira *et al*., 1995). Generally, SSR and SNP markers have higher reproducibility and accuracy compared to the other DNA-based markers.

2.7.1 Non-PCR-Based Molecular Markers

2.7.1.1 Restriction Fragment Length Polymorphism (RFLP) Markers

RFLP analysis was the first technique to be utilized to detect nucleotide variation in DNA sequences, though it is not extensively used today (Botstein *et al*., 1980). It is based on the principle of hybridization – a labelled RFLP probe hybridizes to specific fragment(s) of genomic, chloroplast, or mitochondrial DNA digested by restriction enzymes, which are then separated on agarose or polyacrylamide gels to reveal a distinctive banding pattern unique to a genotype (Navajas & Fenton, 2000). RFLPs exhibit co-dominant inheritance and occur in genomic, chloroplast, and mitochondrial DNA (Weising *et al.*, 2005). RFLP markers have been used to study genetic diversity between domesticated crops and their wild relatives (Devarumath *et al.*, 2002). Matsumoto *et al.* (2002) examined phenylanine ammonia-lyase (PAL) genes by RFLP analysis using PAL-cDNA as a probe to discriminate Assam tea hybrids and Japanese tea cultivars. The alleles of the Japanese cultivars differed greatly from the Korean cultivars, but were similar to Chinese varieties (Matsumoto *et al.*, 2002). RFLP fingerprints are important markers for assessing genetic fidelity in micropropagated tea plants (Devarumath *et al.*, 2002). However, RFLPs show fewer polymorphisms than SSRs and detect limited loci per assay (Navajas & Fenton, 2000). The technique is also time-consuming, labor-intensive, and often requires radioactively labeled probes; hence, it is rarely in use today.

2.7.2 PCR-Based Molecular Markers

2.7.2.1 Random Amplified Polymorphic DNA (RAPD) Markers

This technique is based on PCR amplification of random genomic DNA sequences (Williams *et al.*, 1990). A single 8-19 short primer that anneals at a lower temperature binds to and amplifies several sites on the genome (Williams *et al.*, 1990). RAPD markers were used to evaluate genetic diversity and relationships in 38 Kenyan tea cultivars (Wachira *et al.*, 1995), establish affinities between 28 genotypes of cultivated tea and wild *Camellia* species (Wachira *et al.*, 1997), and map QTLs in 42 tea clones (Kamunya *et al.*, 2010). Genetic variability between these species was found to be significant. In addition, the RAPD fingerprints were able to discriminate between the 38 tea clones that could not be distinguished based on morphological features. Kaundun *et al.* (2000) evaluated genetic diversity of 27 tea accessions drawn from Korean, Taiwanese, and Japanese using RAPD markers. Of the total 50 primers screened, 17 yielded 58 polymorphic and reproducible bands (Kaundun *et al.*, 2000). The study reported highest diversity within the Korean tea relative to Taiwanese and Japanese teas. Although RAPD fingerprints are robust at assessing the genetic fidelity in micropropagated tea plants, they are dominant markers and show limited polymorphism, which makes them less effective in discriminating closely related genotypes (Devaruthmath *et al*., 2002). Additionally, RAPDs display dominant inheritance, limiting homozygote-heterozygote differentiation (Navajas & Fenton, 2000). Other disadvantages include low-level polymorphism detected, limited reproducibility and dominance which prevent heterozygote identification.

2.7.2.2 Amplified Fragment Length Polymorphism (AFLP) Markers

The AFLP technique is considered more reliable and robust in detecting polymorphisms than RAPDs (Vos *et al*., 1995). The underlying principle is the selective PCR amplification of digested DNA fragments. This technique can yield informative fingerprints of genomes with unknown sequences (Weising *et al*., 2005). AFLPs are more efficient in detecting polymorphisms than RFLP and RAPD; hence, can discriminate between closely related genotypes. Paul *et al*. (1997) applied AFLP markers to evaluate genetic variability of 32 tea varieties derived from India and Kenya and could distinguish the three tea germplasms, namely, Assam, China, and Cambod with the Indian Assam genotypes clustering closely with the Kenyan Assam accessions. Although AFLP markers are relatively robust and reliable for population genetic diversity studies, variation in fragment sizes may lead to suboptimal reproducibility, hence limiting the comparability of the banding patterns (Vos *et al*., 1995).

2.7.2.3 Microsatellite Markers

Microsatellite markers, also known as simple sequence repeats (SSRs), are short DNA fragments ($\simeq$ 100bp) that comprise of 2-6bp long motifs repeated in tandem. The number of repeats in microsatellite loci changes extensively during a species' evolutionary history, which accounts for the variation within populations (Putman & Carbone, 2014). Therefore, SSRs are highly abundant per locus, making them excellent tools for genetic diversity studies (Navajas & Fenton, 2000). They also exhibit great reproducibility, display co-dominant inheritance, are multi-allelic, relatively abundant in the genome, and have higher polymorphic information content than RFLPs and RAPDs (Gupta *et al*., 2005; Ellstrand *et al*., 1999). SSR anchored PCR (SSR-PCR) used on 25 tea cultivars clearly distinguished the three clusters of Assam, Cambod, and China genotypes, indicating that the markers could be used to produce genetic fingerprints of tea (Mondal, 2002). Lai *et al*. (2001) also used SSR markers to characterize the genetic relationships in Taiwanese cultivated tea clones and wild types. The Taiwanese wild teas clustered closely with Assam teas than with China teas and the Taiwanese hybrid cultivars. Thus, SSRs are reliable markers for investigating the genetic diversity in tea clones and genetic fidelity in micro-propagated tea plants. Detecting polymorphic loci in genotypes usually entails analyzing the sizes of PCR-amplified fragments (Navajas & Fenton, 2000). The recent vast sequence datasets from expressed sequence tag (EST) projects offer a useful resource for mining and characterizing genic SSRs for diversity studies (Varshney *et al*., 2005). Freeman *et al*. (2004) identified 13 polymorphic SSRs in *C. sinensis* that could be used to study genetic diversity in tea accessions. Ma *et al*. (2010) further reported the development and validation of polymorphism of 74 EST-SSR markers in 45 tea cultivars belonging to seven different varieties. Yao *et al*. (2012) developed and utilized 96 polymorphic EST-SSR markers for analysis of population structure in 450 Chinese tea accessions while Wambulwa *et al*. (2016) isolated and characterized 23 SSR markers that revealed the full extent of the genetic diversity of tea germplasm in East Africa. More recently, 82 SSRs were developed from sequences available in the public databases such as ESTs, GSS, and RNA-seq and validated using 36 tea genotypes (Dubey, 2020). Using novel EST-SSR and validated microsatellites from two previous studies, this study aimed at characterizing gene flow and genetic diversity of interspecific hybrids established in three trials in Kenya.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Experimental Design

The research approach adopted in this study is as summarized in the flow chart below (Figure 3.1). For objective 1, five novel EST-SSR markers were developed from 789 ESTs downloaded from the NCBI database, processed through sequence assembly, removal of contaminating sequences, and SSR motif detection. Subsequently, the five markers and 15 adapted genomic microsatellites were used to screen for polymorphism in four cultivars. For objective 2, genetic diversity studies were conducted in POPGENE v. 1.32 based on 88 tea cultivars and using eight polymorphic SSR markers. For objective 3, gene flow among the 88 cultivars was analyzed using three analyses: population structure, relationship analysis, and parentage analysis.

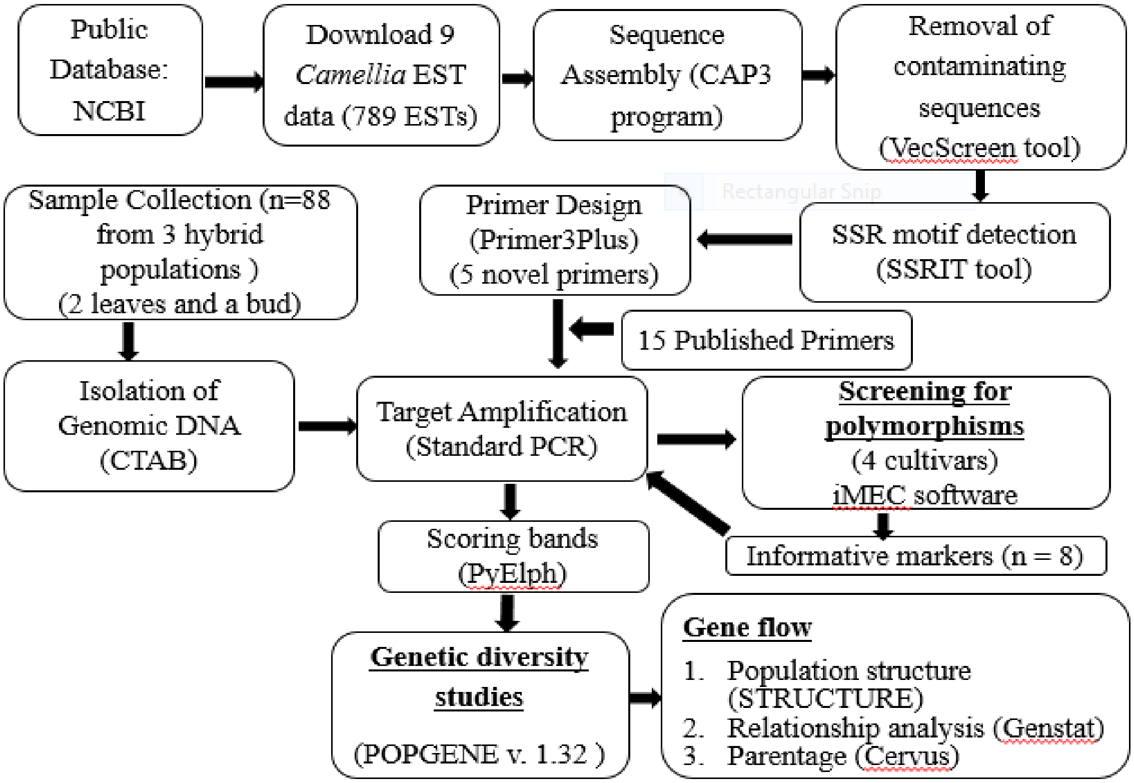

Public Database: NCBI
Download 9 *Camellia* EST data (789 ESTs)
Sequence Assembly (CAP3 program)
Removal of contaminating sequences (VecScreen tool)
Rectangular Snip
Sample Collection (n=88 from 3 hybrid populations) (2 leaves and a bud)
Primer Design (Primer3Plus) (5 novel primers)
SSR motif detection (SSRIT tool)
15 Published Primers
Isolation of Genomic DNA (CTAB)
Target Amplification (Standard PCR)
Screening for polymorphisms (4 cultivars) iMEC software
Scoring bands (PyElph)
Informative markers (n = 8)
Genetic diversity studies (POPGENE v. 1.32)
Gene flow
1. Population structure (STRUCTURE)
2. Relationship analysis (Genstat)
3. Parentage (Cervus)

Figure 3.1: A schematic representation of the experimental design

3.2 Study Site

Tea variety samples were collected from three KALRO-TRI experimental trials established in two sites: Kangaita in Kirinyaga County (0°30′S, 37°17′E, 1548 m.a.s.l) and Timbilil in Kericho County (0º22'S, 35º21'E, 2200 m.a.s.l.) (Figure 3.2). Images of the plots where the interspecific hybrids and wild types were grown are shown in Figure 3.2a-d.

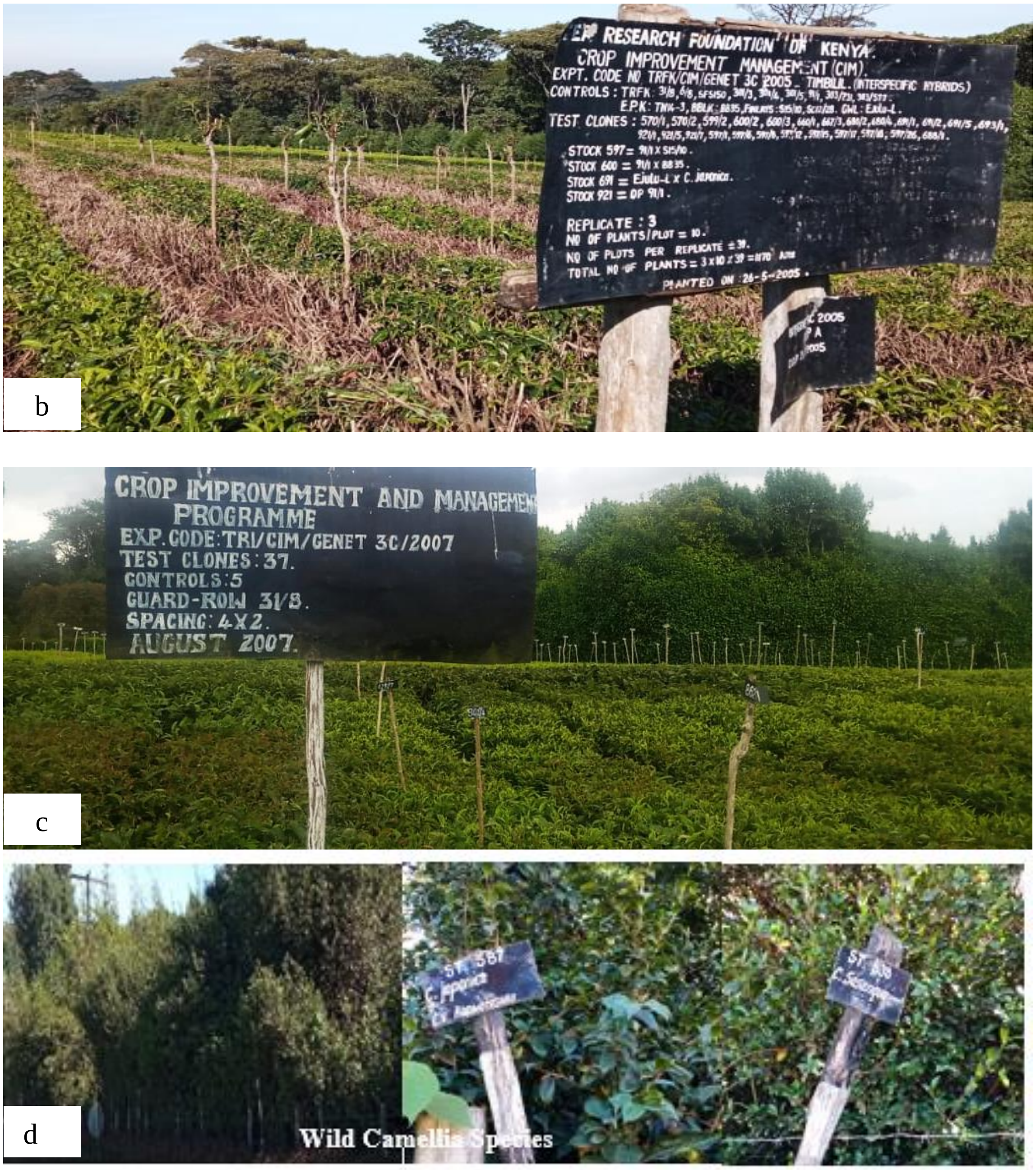

b
RESEARCH FOUNDATION OF KENYA.
CROP IMPROVEMENT MANAGEMENT (CIM).
EXPT. CODE NO TRFK/CIM/GENET 3C 2005 _ TIMBILIL. (INTERSPECIFIC HYBRIDS)
CONTROLS : TRFK:
E.P.K: TN14-3, BBLK: BB35,
TEST CLONES :
STOCK 597 = 91/1 X S15/10.
STOCK 600 = 91/1 X BB35.
STOCK 691 = Ejulu-L x C. japonica.
STOCK 921 = OP 91/1.
REPLICATE : 3
NO OF PLANTS/PLOT = 10.
NO OF PLOTS PER REPLICATE = 39.
TOTAL NO OF PLANTS = 3 x 10 x 39 = 1170
PLANTED ON 26-5-2005.
c
CROP IMPROVEMENT AND MANAGEMEN
PROGRAMME
EXP. CODE: TRI/CIM/GENET 3C/2007
TEST CLONES: 37.
CONTROLS: 5
GUARD-ROW 31/8.
SPACING: 4X2.
AUGUST 2007.
d
ST 387
ST 398
Wild Camellia Species

Figure 3.2a-d: Plots and hybrids for (a) Genet 3c/1999, (b) Genet 3c/2005, (c) Genet 3c/2007 and wild *Camellia* species

The laboratory work was conducted at the Molecular Biology Laboratory of the KALRO-TRI, Kericho County, Kenya.

3.2.1 Establishment and Management of Experimental Plots

3.2.1.1 TRFK/CIM/GENET 3C/1999

The experiment was established at the Kangaita TRI substation, Kirinyaga County in 2002. It comprised of 30 selected tea clones. Of these, 18 are potential interspecific hybrids (colored) between tea and related *Camellia* species selected from seed plantations and 12 are popular ordinary green colored tea clones used for processing of black tea. The trial was established in randomized complete block design with three replicates comprising 10 plants each spaced at 1.22 m by 0.61 m between and within rows, respectively. Each of the three randomized replications had four single-line plots of cultivars that were transplanted from the nursery as 8-12-months old sleeved plants. The purpose of this longitudinal clonal field trial was to evaluate the compare the growth and yield of the hybrids with those of parental clones.

3.2.1.2 TRFK/CIM/GENET 3C/2005

In 2005, 25 potential interspecific hybrids between tea and related *Camellia* species were selected from earlier interspecific crosses (progeny tests in Kangaita) and established among other commercial clones and parental controls in KALRO-TRI, Timbilil Center, Kericho County. The trial has a total 39 different clones included St 570 (TRFK 301/3 x *C. japonica*), St 597 (TRFK 91/1 x AHP S15/10), St 599 (TRFK 91/1 x 301/3), St 600 (TRFK 91/1 x BBK 35), St 660 (TRFK K-purple x AHP SC12/28), St 667 (Taiwan Yamacha 87), St 680 (Vietnam 3), St 691 (GW Ejulu x *C. japonica*), St 693 (TRFK K-purple x TRFK 303/577) and St 921 (TRFK 91/1 OP) established in randomized complete block design with three replicates, as described in section 3.2.1.1.

3.2.1.3 TRFK/CIM/GENET 3C/2007

This trial comprises 38 interspecific hybrid clones derived from crosses St 645 (TRFK 301/4 x TRFK K-Purple), St 862 (TRFK 91/1 x TRFK 301/4), St 688 (TRFK 91/1 x TRFK 303/577), St 845 (TRFK 301/4 x TRFK 91/1) and 4 parental control clones. The trial was established in the TRI Kangaita Centre, Kirinyaga County in 2007 with 15 plants per clonal plot spaced as described in section 3.2.1.1.

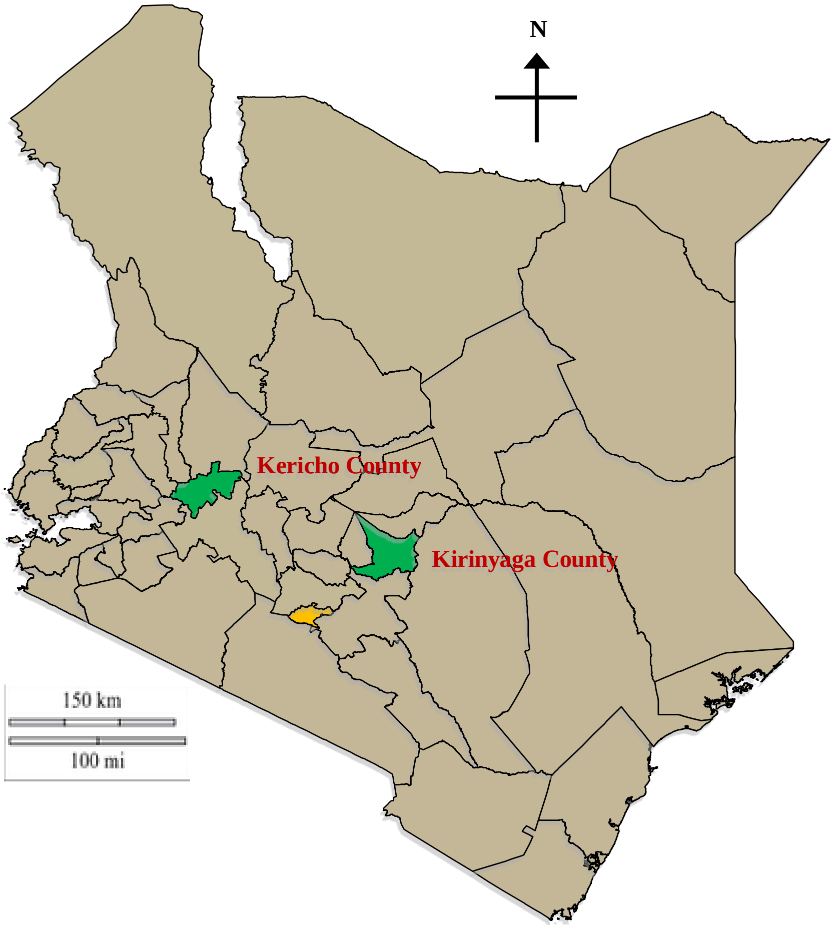


Figure 3.3: A map of Kenya showing the geographical Experimental sites of Kericho County (Timbilil) and Kirinyaga County (RCMRD Geoportal, 2015)

3.3 Sampling Design

Shoots were harvested from the plants established in each plot. Individual experiments contained three replicates of test clones (hybrids) and their controls (parental clones) established in a randomized complete block design. Leaf samples were obtained from all plants belonging to each test clone.

3.4 Sampling of Plant Materials

In this study, the population size was relatively small and shared an uncommon characteristic, i.e., interspecific hybrids, and thus was taken as the sample size to describe the full extent of gene flow and genetic diversity in interspecific tea hybrids. The sample size was 105 comprising all hybrids in Genet 3c/2009 hybrids (n = 18), Genet 3c/2005 (n = 25), Genet 3c/2007 (n = 38), their parents (n = 12) and wild types (n = 12). Fresh tender shoots (two leaves and a bud) were harvested from the hybrid cultivars, their maternal parents, and wild types growing in three experimental trials – two in Timbilil and one in Kangaita, (Figure 3.2). The samples were collected using khaki bags and transported in a cool box to the laboratory, washed in running water, air dried and stored at -20°C for subsequent DNA extraction.

3.5 Genomic DNA Isolation and Purification

Genomic DNA was extracted from the leaf (first flush) samples using modified cetyltrimethylammonium bromide (CTAB) protocol (Kamunya, 2010). About 600g of leaves (equivalent to 2 – 3 shoots) was ground to a fine powder using liquid nitrogen. Then, 4000 µl of 2x CTAB extraction buffer (2% CTAB, 1M Tris-hydrochloric acid pH 8.0, 5M NaCl, 2% polyvinylpolypyrrolidone, 0.5M EDTA, 2% β-mercaptoethanol, and sterile distilled water (SDW)) pre-heated at 65ºC for 60 min was added. The extract was transferred to a 15ml centrifuge tube and incubated in a water bath at 65ºC for 30 min. Thereafter, 700µl of chloroform:isoamylalcohol (CIA) (24:1) was added then the mixture vortexed and centrifuged at 7000 rpm for 15 min. The supernatant was transferred to a fresh tube and 700 µl of ice-cold isopropanol was added. After gentle inversion of the tube, the mixture was centrifuged at 5000 rpm for 5 min and the supernatant discarded. Ice-cold ethanol (70%) was added then the tube centrifuged (5000 rpm for 5 min) and the supernatant discarded leaving the pellet. The pellet was air-dried, then dissolved in 1000µl SDW and 2µl RNAse added before overnight incubation in a water bath at 55^{o}C.

After overnight incubation, the extracted DNA was purified by the addition of 1ml CIA and the mixture shaken for 15 min. Subsequently, this was centrifuged at 7800 rpm for 15 min and the aqueous phase transferred to a new microcentrifuge tube and 200 µl of NaCl-TE added to the old tube. The tube was then shaken for 15 min and centrifuged at 7800 rpm for 15 min. The aqueous phase was transferred to the microcentrifuge tube and 800 µl ice-cold isopropanol alcohol added. The mixture was centrifuged at 5000 rpm for 5 min, then the supernatant discarded and the pellet rinsed with 1000 µl ice-cold ethanol (70%) before the pellet was air-dried in a lamina airflow. Dry pellet was dissolved in 100 µl sterile distilled water and kept at 4^{o}C. DNA quality and quantity were assessed using Nanodrop spectrophotometry (Thermo Scientific NanoDrop 2000 UV-Vis Spectrophotometer), while integrity was checked using 1.5% agarose gel electrophoresis.

3.6 Data Mining and Processing of EST-SSRs

A total of 789 Expressed Sequence Tags (ESTs) belonging to *C. japonica* (519 ESTs), *C. taliensis* (67 ESTs), *C. brevistyla* (59 ESTs), *C. chrysantha* (45 ESTs), *C. furfuracea* (37 ESTs), *C. sasanqua* (28 ESTs), *C. kissi* (19 ESTs), *C. irrawadiensis* (9 ESTs), and *C. assimilis* (6 ESTs) were downloaded from the GenBank of the NCBI (http://www.ncbi.nlm.nih.gov/) in their FASTA format on 20th September 2019. After removal of redundancy in the sequences using a sequence assembly software, CAP 3 program, with default parameter values (i.e. base quality cutoff for clipping = 12, overlap length cutoff = 30, overlap percentage identity cutoff = 75, overlap similarity score cutoff = 500 and minimum number of good reads at clip position = 2) (Huang & Madan, 1999), 440 non-redundant unigenes (NR) (80 contigs and 360 singletons) were generated. Further, contaminating sequences such as adapters, linkers, PCR primers, and vector sequences were removed by screening the NR sequences against the UniVec database (ftp: //ftp.ncbi.nih.gov/pub/ UniVec/) using the VecScreen tool (http://www.ncbi.nlm.nih.gov/tools/vecscreen/) set at Expect min match = 10 and

Percentage identity min = 10. Subsequently, polyA/T tails were trimmed from the ESTs with the EST_trimmer.pl script (http: //pgrc.ipk-gatersleben.de /misa /download /est_trimmer.pl) until no low complexity segments, (A)n or (T)n, remained on either the 3' or 5' end.

3.7 Identification of SSR Motifs and Primer Design

To detect SSR motifs, the non-redundant EST datasets of the nine *Camellia* spp. were separately processed using the Sequence Repeat Identification Tool, SSRIT (Temnykh *et al*., 2001). The criteria used were as follows: maximum motif length = decameter (10 identical and repetitive nucleotides); minimum number of repeats allowed = 3. The microsatellites were classified into Class I (≥ 20 nucleotides) and Class II (12 to ≤ 20 nucleotides) and used in primer design.

PCR primers were designed using Primer3Plus software based on the regions flanking each SSR motif (Rozen & Skaletsky, 2000; SantaLucia, 2007). The design parameters were set as follows: primer length 18-27bp, optimum 20bp; annealing temperature (Tm) minimum 57℃, maximum 63℃ and optimum 60℃; %GC content min 40, max 60, and optimum 50; maximum Tm difference between sense and antisense primer 2℃; and amplicon size range from 125 to 300 bp (SantaLucia, 2007). Additionally, the number to return was set at 5, Max 3' Self complementarity was set at 0 and increased by 1 if no primers returned), and Max Poly X set at 1 (an increment of 1 if no primers are returned). Fourteen functional EST-SSR primers were designed and synthesized by Inqaba Biotec, South Africa with five of them matching the criteria described by Zhang *et al*. (2016) (Appendix 1).

3.8 Screening and Validation of SSR Markers

A total of 20 SSR primers comprising five novel and 15 adapted from published work – 10 from Wambulwa *et al*. (2016) and five from Freeman *et al*. (2004) – (Appendix 1) were screened for polymorphism using a subset of the cultivars (n=4). The cultivars consisted of TRFK 570/2 (progeny of cross TRFK 301/3 ♀ and *C. japonica* ♂), TRFK 688/1 (progeny of cross *C. irrawadiensis* ♀ and TRFK 303/577 ♂), TRFK 83/1 (clonal bush obtained from Kapchomo Estate, EPK Nandi in 1966), and TRFK 6/8 (commercial standard cultivar in processing high quality black tea).

PCR amplifications were performed in 10 µL reaction volume (Appendix 2) using a thermal cycler (TC-5000, Techne Inc., Thermo Scientific) each consisting 40 ng genomic DNA, 0.2mM dNTPs, 0.5U *Taq* polymerase, 2 mM $MgCl_2$, 0.5 µM of each primer (Forward and Reverse), 1× PCR buffer (100 mM Tris-HCl, 500 mM KCl; pH 8.3), and two drops of mineral oil to prevent sample evaporation. Standard PCR was run with a specific SSR program: initial denaturation for 4 min at 94°C, followed by 35 cycles of 94°C for 30s, 55°C for 1 min, 72°C for 30s, and a final extension of 7 min at 72°C (Appendix 3).

On the basis of polymorphic information content, discriminating power, and number of polymorphic bands, eight polymorphic SSR primers were selected as ideal for studying genetic diversity in the cultivars used in this study (n=88). PCR amplifications for all the genotypes were done using the conditions reported above.

The PCR products were resolved on 1.5% agarose gel run in 1x TBE buffer for 180 min at 150 V (Bio-Rad model 200/2.0 power supply and wide mini-sub cell GT horizontal electrophoresis system, Bio-Rad laboratories, Inc., USA) and stained with Ethidium Bromide (etBr) (0.5 µg/ml) solution for 40 min. In order to determine the molecular size of the amplified products, each gel was loaded with 6µl of 50bp DNA size standard (Inqaba, South Africa). Finally, the gels were visualized under UV light at 312nm in a gel documentation system (UVP PhotoDoc-it^{TM} imaging system + Benchtop Variable Transilluminator Upland, CA, USA).

3.9 Data Analysis

Binary data (1=band present, 0=band absent) were generated from gel images of amplified fragments of SSRs using PyElph software (Pavel & Vasile, 2012). First, each loaded gel image was rotated so that the wells are at the top of the image view. Subsequently, the lanes were detected automatically with parameters: lane width=40 and width deviation=25%. Bands were then detected using the following parameters: filter threshold=38, filter width=3, and filter passes=10. A band matching operation was then used to cluster bands of similar size (a distance parameter of 2%) to produce a matrix data for all populations and loci that were exported to MS Excel for analysis.

To assess the informativeness of the markers, an Online Marker Efficiency Calculator (iMEC) was used to compute key indices of polymorphism: the polymorphic information content (*PIC*) and discriminating power (*D*) (Amiryousefi *et al*., 2018). The *PIC* of each primer-pair was estimated using the following formula;

$$PIC = 1 - \left(\sum_{i=1}^{n} p_i{}^2\right) - \sum_{i=1}^{n-1}\sum_{j=1+1}^{n}\left(2p_i{}^2 p_j{}^2\right)$$

Where *pi* and *pj* are the distribution frequencies of the *i*-th and *j*-th alleles in the population, whereas *n* denotes the number of alleles identified by a marker (Amiryousefi *et al*., 2018). *PIC* indicates the discriminating power of a marker based on allele distribution frequency and the number per locus in the genotypes being studied (Nagy *et al*., 2012). Co-dominant markers such as microsatellites SSR markers with a *PIC* value of ≥0.3 can detect moderate to high genetic diversity in a population (Botstein *et al*., 1980; Amiryousefi *et al*., 2018).

The discriminating power (D) of each marker was computed using the following formula;

$$D = 1 - C_i = 1 - \sum_{i=1}^{I} P_i \frac{(NP_i - 1)}{N - 1}$$

Where *I* is the total number of genotypes (banding patterns) produced by a marker, *Pi* denotes the frequency of *ith* genotype of the *jth* primer, *N* represents the number of individuals tested, and *Ci* the confusion probability of *jth* SSR, which is the likelihood that any two individuals selected randomly from a sample possess a similar banding pattern (Nei & Li, 1979).

The genetic diversity of the 88 *Camellia* genotypes was studied with the POPGENE version 1.32 program (Yeh *et al.*, 1999), which measured the following parameters: number of polymorphic bands and percentage of polymorphic bands, observed (Na) and effective number of alleles (Ne) per locus, Shannon information index (I), observed heterozygosity (Ho), expected heterozygosity (He), and gene flow (Nm). Nm was estimated as: $Nm = 0.25(1-Fst) / Fst$. Multi-population and single-population genetic diversity indices were computed with this software. The same diversity indices were computed for hybrid families within these populations. F-statistics that measure the genetic structure of the population were also computed using POPGENE version 1.32 software.

The population structure of the *Camellia* cultivars was analyzed using the Bayesian model-based clustering algorithms in the program Structure v. 2.3.4 (Pritchard *et al.*, 2000). The parameter set included 10 runs (ranging from 2 to 9) and 10^5 Markov Chain Monte Carlo replicates after allowing a burn-in period of 10^5 interactions for each group. Correlated allele frequencies in the admixture model were used as the individuals were assumed to have mixed ancestry.

The analysis of molecular variance (AMOVA) between and within the five populations was performed with the GenAlEx 6.5 software (Peakall & Smouse, 2006) without grouping the populations into geographical regions based on fragment size data. Parentage analysis was performed using Cervus 3.0.7, a computer program that assigns parents to offspring based on genetic markers and involves two assumptions: species are diploid and markers are in linkage equilibrium (Kalinowski *et al*., 2007). Parentage analysis was done to estimate the resolving power of codominant loci given their allele frequencies and estimate critical values of the log-likelihood statistics LOD, so that the confidence of parentage assignments can be evaluated statistically (Kalinowski *et al*., 2007). Two types of parentage analyses were performed: maternity and paternity analysis with known maternal parents.

Further relationship analysis involved constructing a dendrogram by the neighbor joining (NJ) method from the Genstat (15th edition) program based on Jaccard's similarity coefficient. Genetic structure analysis of the population was conducted using STRUCTURE 2.3.4 with 10 runs and 10^5 Markov Chain Monte Carlo repetitions allowing a burn-in period of 10^5 iterations for each group (K) from 2 to 8. The optimal K value was determined based on the estimated probability of K (*Ln* P(D) that captures the structure of the data (Pritchard *et al*., 2000).

CHAPTER FOUR

RESULTS

4.1 Description of Samples

Of the sampled 105 accessions, only 88 tea samples gave quality DNA that were subsequently used in the present study. High concentration of polyphenols and phenolic compounds in the leaves of some of these hybrids could account for the low quality DNA obtained (Graham, 1992). The 88 genotypes used represented the wild type, interspecific test varieties and their parental controls. A list of the varieties used in the study and their location details are summarized in Table 4.1.

4.2 DNA Quality and Quantity

The sample concentration of genomic DNA (gDNA) ranged from 330.80 ng/µl to 15,711.80 ng/µl (M=2940.34, SD = 2027.46) (Table 4.2) Variations in the leaf sample used in DNA extraction and loss of DNA during phase separation could account for the large variance in the DNA yield. The determination of DNA purity based on the absorbance ratio of 260nm/280nm showed a maximum ratio of 2.01 and a minimum of 1.53 (M= 1.79, SD= 0.098) with samples measuring ≥1.80 considered to be of high quality and purity and subsequently used in PCR.

Table 4.1: List of varieties and their location details used to study genetic diversity in *Camellia* spp. using SSR markers

Wild type (Timbilil)	Genet 3c/1999 (Kangaita)	Genet 3c/2007 (Kangaita)	Genet 3c/2005 (Timbilil)	Parents (Timbilil)
C. irrawandiensis	TRFK 31/38	TRFK645/14	TRFK 921/5	TRFK 6/8
C. kissi	TRFK 31/36	TRFK 645/6	TRFK 921/1	EPK TN14-3
C. oleifera	TRFK31/35	TRFK 645/5	TRFK 691/1	TRFK 301/2
C. brevistyla	TRFK 31/34	TRFK 862/5	TRFK 680/2	TRFK 31/8
C. sasanqua	TRFK 31/33	TRFK 862/4	TRFK 667/3	AHP SC12/28
C. japonica	TRFK 31/32	TRFK 862/3	TRFK 660/1	BBK BB35
	TRFK 31/11	TRFK 862/1	TRFK 600/3	AHP S15/10
	TRFK 301/1	TRFK 845/6	TRFK 599/2	TRFK 301/3
	TRFK 14/1	TRFK 845/5	TRFK 597/26	GW Ejulu-L
	TRFK 91/2	TRFK 845/4	TRFK 597/17	TRFK K-purple
	TRFK 73/5	TRFK 845/3	TRFK 597/15	TRFK 301/4
	TRFK 73/4	TRFK 845/2	TRFK 597/12	TRFK 303/577
	TRFK 73/3	TRFK 845/1	TRFK 597/8	
	TRFK 73/2	TRFK 688/19	TRFK 597/1	
	TRFK 73/1	TRFK 688/18	TRFK 570/1	
	TRFK 306/4	TRFK 688/13	TRFK 691/2	
	TRFK 306/3	TRFK 688/12	TRFK 688/1	
	TRFK 306/2	TRFK 688/11	TRFK 570/2	
	TRFK 306/1	TRFK 688/10		
	TRFK 83/1	TRFK 688/7		
		TRFK 688/6		
		TRFK 688/4		
		TRFK 688/1		
		TRFK 862/20		
		TRFK 862/22		
		TRFK 688/15		
		TRFK 862/16		
		TRFK 862/14		
		TRFK 862/11		
		TRFK 862/7		
		TRFK 862/6		
		TRFK 862/9		
Total **6**	**20**	**32**	**18**	**12**

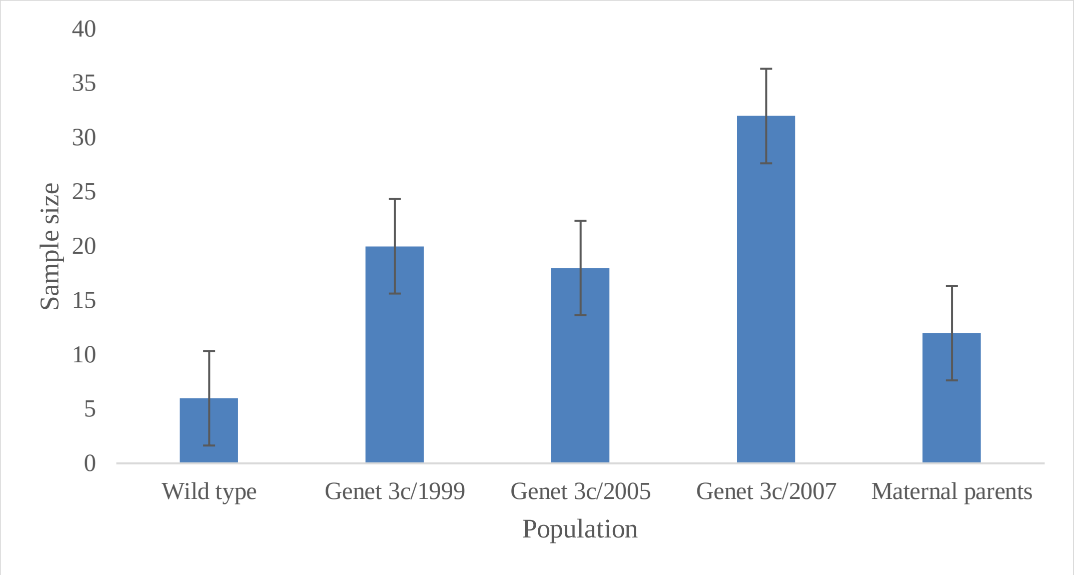


Figure 4.1: Error-bar chart of the number of varieties from each population used in the study

Table 4.2: Concentration and absorbance of isolated genomic DNA used in the study

Statistic	**DNA Quantity (ng/µl)**	**Absorbance at 260nm**	**Absorbance at 280nm**	**DNA Purity (260nm/280nm)**
Range	15381.00	307.62	176.01	0.48
Minimum	330.80	6.62	3.88	1.53
Maximum	15711.80	314.24	179.89	2.01
Mean	2940.34	59.08	33.06	1.79
Std. error	214.91	4.32	2.47	0.01
Std. deviation	2027.46	40.77	23.27	0.10
Variance	4110573.76	1662.01	541.59	0.01
Skewness	3.02	2.95	3.07	-0.38
Kurtosis	17.09	16.60	17.28	0.13

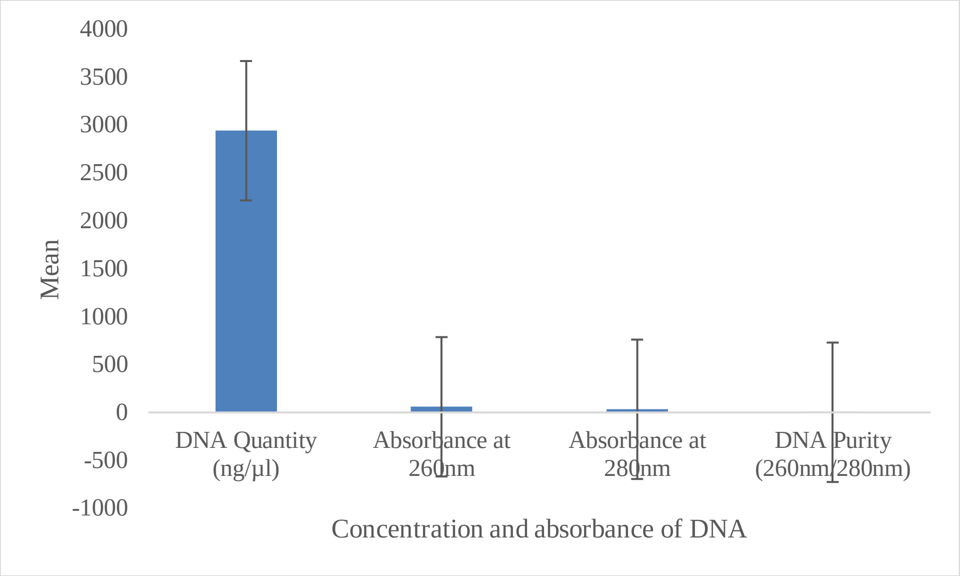


Figure 4.2: Error-bar chart of concentration and absorbance of DNA

4.3 Development of EST-SSR Primers

A total of 1331 potential SSR repeats (221 in 80 contigs and 1,110 in 360 singletons) were identified by the SSRIT tool from the 789 ESTs belonging to the nine *Camellia* spp. that were downloaded from the NCBI database. This represented 18.6% of the unigenes with microsatellite motifs. Di-nucleotides were the most abundant repeat motif with 836 (62.90%) loci followed by tri-nucleotides with 456 (34.31%) loci (Figure 4.1). The remaining loci, consisting of tetra-, penta-, hexa-, and hepta-nucleotides, collectively accounted for 2.78% (37 loci). Mononucleotides were omitted since they could have resulted from sequencing errors (Taheri *et al*., 2018). On average, the maximum number of repeats (CT) found in one unigene were 14. Di-nucleotide repeats of the (TA)n, (AT)n, and (AG)n type were the most abundant microsatellites at 30.8% followed by (TG)n, (GA)n, and (TC)n at 20.8% (Figure 4.2). Among tri-nucleotides, (CCA)n, (ACC)n, (GAA)n and (CAC)n repeats were the most prevalent, cumulatively occurring in 28 sequences.

In total, 170 microsatellites comprising 39 Class I and 131 Class II types were detected. However, only fourteen returned functional EST-SSR markers based on Primer3Plus optimal design parameters. Finally, 5 polymorphic primer pairs flanking these EST-SSRs with fairly similar Tm and %GC, minimal or no secondary structure (primer-dimers), annealing temperature ~60^0C, and GC content of less than 50% markers specific to tri-nucleotide SSRs were randomly selected on the basis that they were likely to be maintained in related species due to triplet codon (Tessier *et al*., 1999; Zhang *et al*., 2016). The primers were synthesized and tested alongside fifteen adapted primers whose polymorphism had been established (Freeman *et al*., 2004; Wambulwa *et al*., 2016).

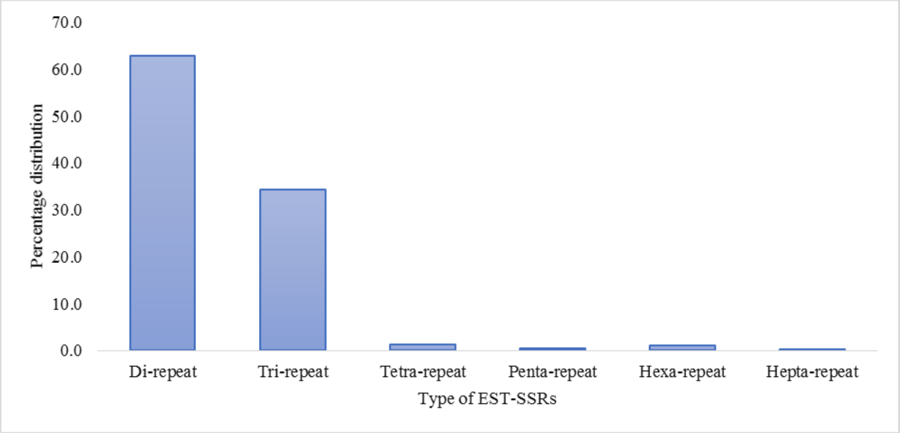


Figure 4.3: EST-SSRs repeat motif percentage in unigenes belonging to *Camellia* spp.

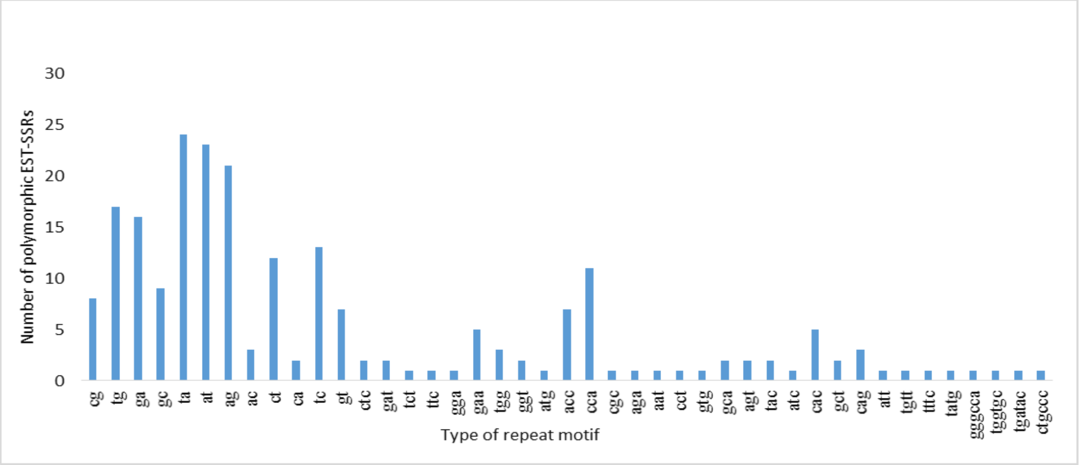


Figure 4.4: Repeat motif type distribution of polymorphic EST-SSRs belonging to *Camellia* spp.

4.4 Polymorphisms and Discriminating Power of SSR Markers

To validate the polymorphisms of the 20 SSR markers, PCR-based genotyping was performed using three randomly selected interspecific hybrids and one commercial/control cultivar. Of these primers, 16 produced PCR amplicons that were separated by size in all the three interspecific hybrids, while 4 amplified in only two interspecific genotypes (TRFK 570/2 and TRFK 688/1), and 14 in the commercial intraspecific cultivar, TRFK 6/8 (Figure 4.3). A total of 85 bands was identified (Table 4.3) using the 20 primers with the number of bands per locus ranging from 1 (Camsin M4) to 22 (Camjap A1) with an average of 4.25 (SD = 4.59). The number of bands were equivalent to the number of SSR alleles or allelic frequency at each locus. SSR markers with a high allele frequency were Camjap A1, TM 134, Camjap A4, and A47, at 22, 10, 8, and 5, respectively (Table 4.3). The size of the amplified alleles at all loci varied between 50 bp and 1500 bp, with a mean of 491.78 bp (SD = 439.33).

A total of 28 polymorphic SSR alleles were produced by 14 SSR primers (Camjap A1, Camjap A2, Camjap 4, TM 134, TM 179, TM 197, TM 203, TM 51, A37, A47, Camsin M1, Camsin M3, Camsin M3, and Camsin M5), accounting for 32.94% polymorphism in the four cultivars (Table 4.3). The *PIC* value ranged from 0.00 (Camsin M4) to 0.53 (A37), with a mean of 0.26 per marker (SD = 0.13). The mean *PIC* value for genomic (adapted) microsatellites was 0.26 compared to 0.28 for EST-SSR (novel) markers. On the basis of *PIC* values, two markers – Camjap A4 and A47 – were highly informative (*PIC* ≥ 0.5), whereas Camjap A1, TM134, and A37 were relatively informative (*PIC* ≥ 0.4). However, whereas the correlation between *PIC* values and the number of alleles detected was significant for the highly informative markers (r = 1.0, p = 0.01), it was non-significant for the relatively informative markers (r = 0.732, p = 0.24).

The discriminating power (D) of the 20 markers averaged 0.142 (SD = 0.20). Two markers (Camjap A4 and A47) showed a higher discriminating power D ≥ 0.5 (M = 0.23, SD = 0.26) (Table 4.3). On the basis of *PIC* (≥ 0.20), discriminating power (D ≥ 0.10), and number of polymorphic bands (≥ 1), a set of eight polymorphic SSR primers (Camjap A1, Camjap A4, TM 134, TM 58, A37, A47, Camsin M2, and Camsin M5) which had mean *PIC* and D values of 0.40 and 0.30, respectively indicating efficient ability to discriminate hybrid cultivars (Table 4.4), were selected as ideal for studying genetic diversity in interspecific tea hybrids. On average, the number of alleles and the number of polymorphic bands detected by the eight markers were 6.9 and 2.8 per locus, respectively (Table 4.3).

Table 4.3: Characteristics of the SSR primers used to screen for polymorphisms in four *Camellia* genotypes

Primer #	**Primer's Code**	**Allele No.**	**Size range (bp)**		**No. of polymorphic bands**	***PIC* value**	**Discriminating power (D)**
			Min.	**Max.**			
1	Camjap A1	22	50	1500	12	0.375	0.753
2	Camtal A1	3	50	200	0	0.157	0.000
3	Camjap A2	4	50	450	1	0.240	0.079
4	Camjap A3	3	50	200	0	0.157	0.000
5	Camjap A4	8	50	400	3	0.449	0.151
6	TM 134	10	125	1350	1	0.372	0.698
7	TM 179	3	200	850	1	0.337	0.130
8	TM 197	3	100	200	1	0.337	0.056
9	TM 203	2	150	200	1	0.190	0.056
10	TM 51	2	125	200	1	0.190	0.056
11	TM 58	2	200	225	1	0.194	0.074
12	TUGMS 2-135	2	200	225	0	0.178	0.000
13	TUGMS 2-143	2	200	250	0	0.194	0.074
14	A37	4	50	750	1	0.446	0.204
15	A47	5	50	750	2	0.527	0.222
16	Camsin M1	2	50	250	1	0.190	0.056
17	Camsin M2	3	200	250	1	0.337	0.130
18	Camsin M3	2	150	200	1	0.190	0.056
19	Camsin M4	1	300	-	0	0.000	0.000
20	Camsin M5	2	125	150	1	0.190	0.056
Total	**-**	**85**	**-**	**-**	**28**	**-**	**-**
Average	**-**	**4.25**	**-**	**-**	**0.9**	**0.262**	**0.142**
SD	**-**	**4.59**	**-**	**-**	**18**	**0.125**	**0.204**

Table 4.4: Characteristics of SSR primers showing informativeness on four *Camellia* genotypes

Primer #	Primer's Code	Allele No.	No. of polymorphic bands	*PIC* value	Discriminating power (D)
1	Camjap A1	22	12	0.38	0.75
2	Camjap A4	8	3	0.45	0.15
3	TM 134	10	1	0.37	0.70
4	TM58	2	1	0.19	0.10
5	A37	4	1	0.45	0.20
6	A47	5	2	0.53	0.22
7	Camsin M2	2	1	0.34	0.13
8	Camsin M5	2	1	0.19	0.10
Total	-	55	22	-	-
Average	-	6.9	2.8	0.40	0.30

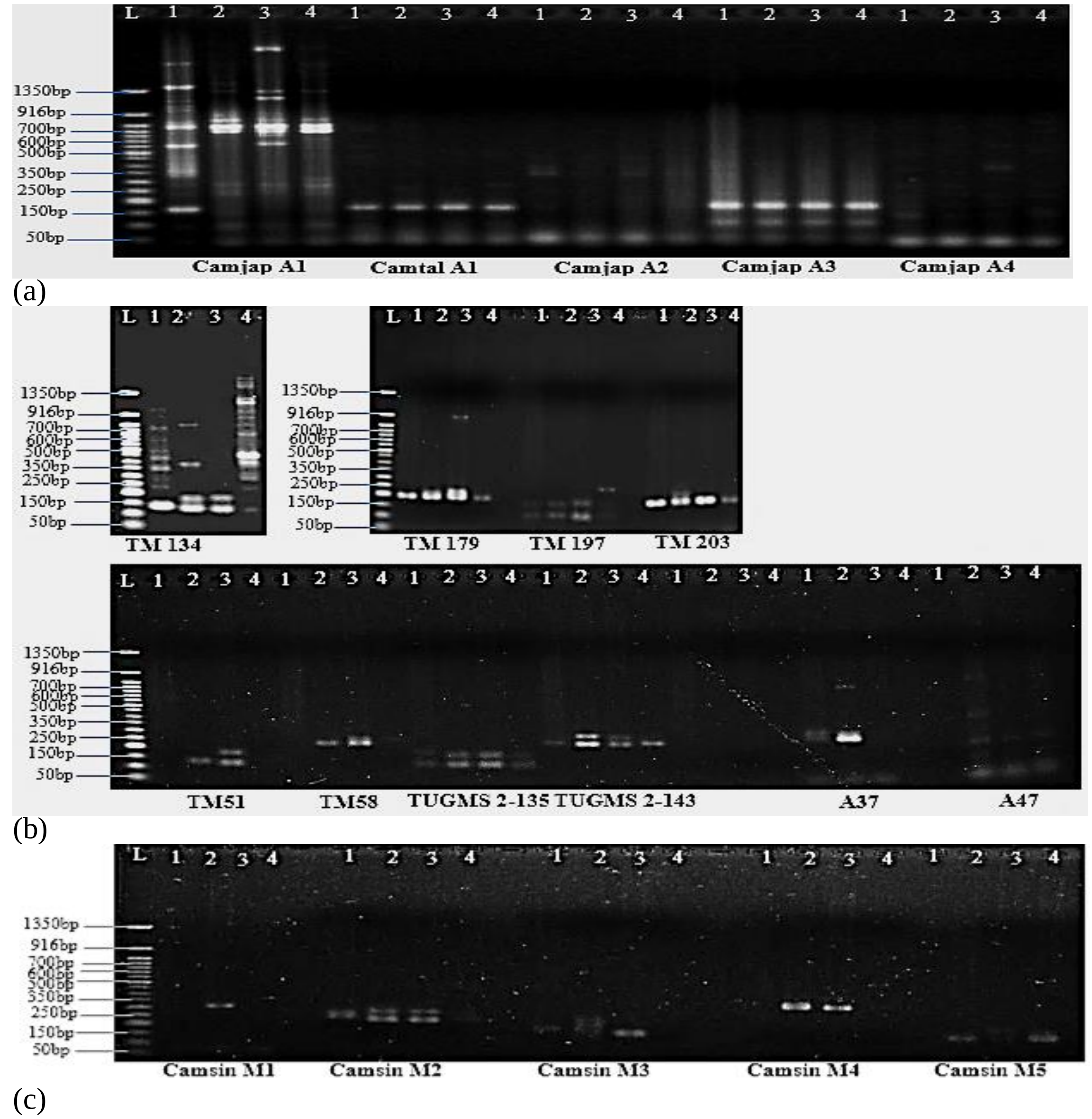


Figure 4.5: SSR marker profiles (a) Camjap A1, Camtal A1, Camjap A2, Camjap A3, Camjap A4, (b) TM 134, TM 179, TM 197, TM 203, TM 51, TM 58, TUGMS 2-135, TUGMS 2-143, A37, A47, and (c) Camsin M1, Camsin M2, Camsin M3, Camsin M4, and Camsin M5) of intraspecific hybrid 1 (6/8 – positive control) and interspecific hybrids 2 (570/2), 3 (688/1), and 4 (83/1) on ethidium bromide- stained 2% agarose gel using 50 bp DNA size marker (L) (Inqaba Biotech, South Africa).

4.5 Genetic Diversity of Interspecific Tea Hybrids

4.5.1 SSR Variation and Genetic Diversity

Gel images of amplified fragments separated on 1.5% agarose were used to score clear bands (Figure 4.4 and Figure 4.5). A total of 2135 bands was scored at the eight loci among the 88 *Camellia* accessions and a matrix (1= band present, 0 = band absent) was generated (Appendix 2). The indices of genetic variation between and within the *Camellia* populations based on SSR markers are shown in Table 4.3 and Table 4.4, respectively. Among the populations studied (wild type, interspecific hybrids (half- and full-sibs), and parental population), little differences were observed in most genetic diversity parameters such as the effective number of alleles (Ne), observed heterozygosity (Ho) and expected heterozygosity (He) and Shannon Information Index (I). Although most of the fragments (at four loci) were monomorphic in the tested genotypes (Camsin M2, Camjap A1, Camjap A4 and A37), 9 polymorphic bands were identified with the highest having 4 at locus Camsin M5. Allelic diversity (mean number of observed alleles per locus) in the 88 genotypes was 2.0. The number of effective alleles in all the tested genotypes ranged from 1.39 (Camsin M5) to 1.65 (TM 51) with a mean of 1.52 (Table 4.5). Multi-population variation characterized using Shannon's information index (I) among the eight loci, ranged from 0.452 at locus Camsin M5 to 0.584 at locus TM 51, with a mean of 0.522 among the 88 genotypes.

Within-population variation was characterized using Shannon's information index (I). It ranged from 0.450 in wild-type population to 0.686 in Genet 3c/2007. Populations Genet 3c/1999 and parents also scored higher I values >0.5 (Table 4.5). The effective number of alleles was highest in Genet 3c/2007 (Ne = 1.973) and lowest in the wild-type population (Ne = 1.43).

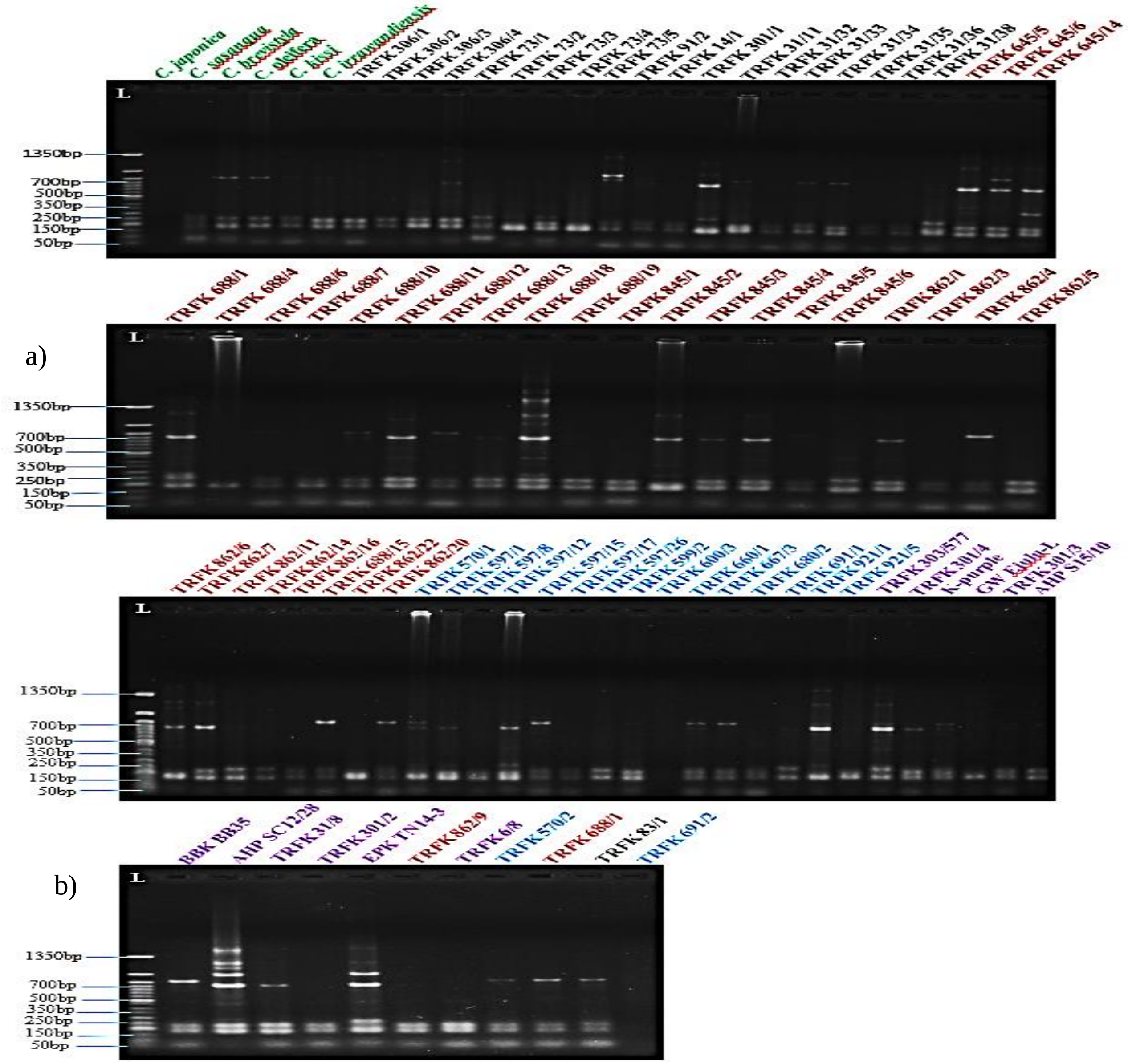
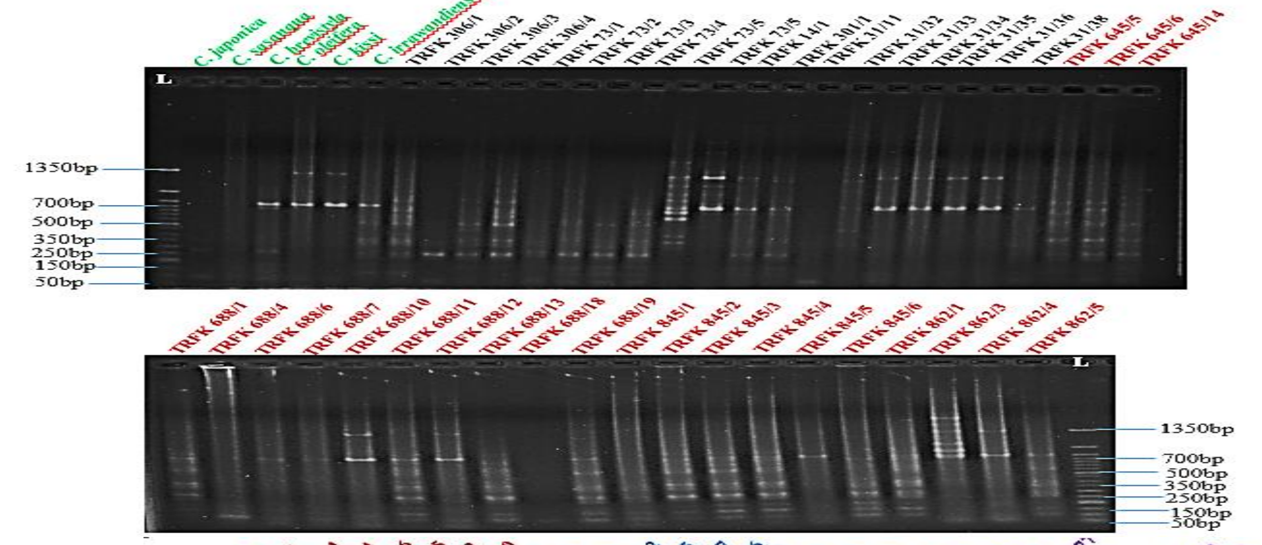


Figure 4.6: Representative gel pictures (a & b) showing bands for amplified *Camellia* cultivars of wild type (in green) and hybrids from Genet 3c/1999 (in black), Genet 3c/2007 (in red), Genet 3c/2005 (in blue), and parents – positive controls (in purple)

with primer Camsin M5. Two genotypes (*C. japonica* and 691/2 were not amplified by this primer. L: 50 bp DNA size marker (Inqaba Biotech, South Africa).



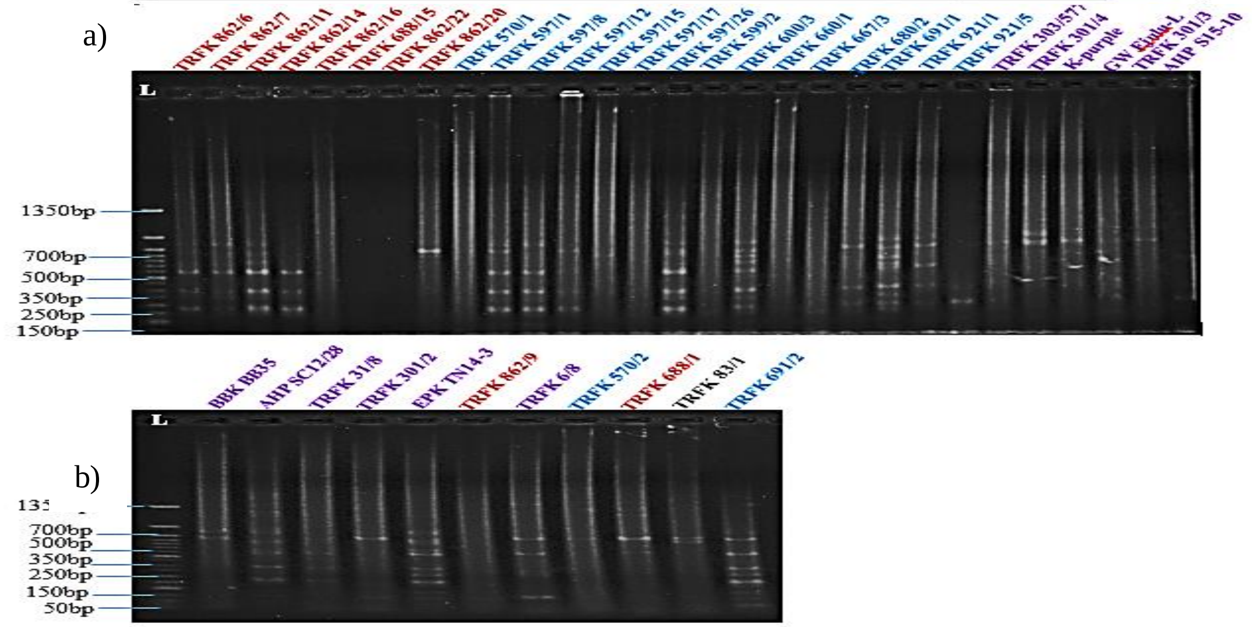


Figure 4.7: Representative gel pictures (a & b) showing bands for amplified *Camellia* cultivars of wild type (in green), Genet 3c/1999 hybrids (in black), Genet 3c/2007 hybrids (in red), Genet 3c/2005 hybrids (in blue), and parents – positive controls (in purple) with primer Camjap A1. Two genotypes (688/15 and 862/22) did not amplify with this primer. L: 50 bp DNA size marker ((Inqaba Biotech, South Africa).

4.5.2 Genetic Differentiation of Interspecific Tea Hybrids

The genetic differentiation (F_{st}) per locus ranged from 0.0115 (TM 134) to 0.1656 (Camjap A1) with an average of 0.0661 alleles per locus (SD = 0.0436), suggesting low genetic differentiation among the populations (Table 4.6). Additionally, based on Shannon information index, the average genetic diversity within populations was not significantly different from that found among populations (I = 0.5181 vs. 0.5216, respectively) ($p < 0.05$). Higher values generally indicate high diversity levels, implying that GENET 3c/2007 (I = 0.6862) is the most diverse population, while wild teas (I = 0.4105) are the least diverse.

Gene flow was highest (implying lower genetic differentiation among groups) at TM 134 (Nm = 21.406) and lowest at locus Camsin M5 (Nm = 1.2593) with an average of 6.5624 (SD=6.3670). AMOVA analysis revealed that 97% of the molecular variation in the tested *Camellia* genotypes existed within individual genotypes and 3% among populations (Table 4.7), probably due to the higher rates of gene flow (Nm = 6.5624) between the populations. Among the three molecular variance indices (F_{is}, F_{st}, and F_{it}), only F_{st} was highly significant ($p < 0.001$).

Table 4.5: Genetic diversity at 8 SSR loci characterized using Shannon's Information Index

Locus	***NPB***	***PPB***	***Na***	***Ne***	**I**	***Ho***	***He***	***Nm***	***Fst***
Camsin M5	4	1.10	2.0000	1.3876	0.4524	0.0000	0.2803	1.2593	0.1656
Camsin M2	0	0.00	2.0000	1.4706	0.5004	0.0000	0.3209	12.048	0.0203
Camjap A1	0	0.00	2.0000	1.4588	0.4940	0.0000	0.3152	3.3674	0.0691
Camjap A4	0	0.00	2.0000	1.4470	0.4875	0.0000	0.3096	2.9948	0.0770
TM 51	1	0.04	2.0000	1.6522	0.5838	0.0000	0.3971	3.8953	0.0603
TM 134	2	1.42	2.0000	1.6225	0.5718	0.0000	0.3837	21.406	0.0115
A 37	0	0.00	2.0000	1.6000	0.5623	0.0000	0.3750	3.6020	0.0649
A 47	2	0.68	2.0000	1.5091	0.5203	0.0000	0.3373	3.9266	0.0599
Total	9	-	-	-	-	-	-	-	-
Mean	1.125		2.0000	1.5185	0.5216	0	0.3399	6.5624	0.0661
St. Dev			0.0000	0.0891	0.0436	0	0.0385	6.3670	0.0436

NB: *NBP* = number of polymorphic bands, *PPB* = percentage of polymorphic bands, *Na* = number of observed alleles, *Ne* = number of effective alleles, *I* = Shannon's information index, *Ho* = observed heterozygosity, *He* = expected heterozygosity, *Nm* = gene flow, *Fst* = coefficient of genetic differentiation.

Genetic diversity analysis was done for 11 hybrid families each represented with at least two members. Moderate to high diversity was revealed by St 31 comprising TRFK 31/11, TRFK 31/32, TRFK 31/33, TRFK 31/34, TRFK 31/35, TRFK 31/36, and TRFK 31/38 showing the least diverse (*I* = 0.36) (Table 4.8), whereas St 645 with TRFK 645/14, TRFK 645/6, and TRFK 645/5 had the highest diversity (*I* = 0.64). Genetic diversity was also relatively high in two other families: St 570 (*I* = 0.45) represented by TRFK 570/1 and TRFK 570/2 and St 688 (*I* = 0.45) represented by TRFK 688/19, TRFK 688/18, TRFK 688/13, TRFK 688/12, TRFK 688/11, TRFK 688/10, TRFK 688/7, TRFK 688/6, TRFK 688/4, and TRFK 688/1.

Table 4.6: Single-population genetic diversity studies

Population										
	Wild-type (n=6)		**3c/1999 (n=20)**		**3c/2007 (n=32)**		**3c/2005 (n=18)**		**Parents (n=12)**	
Locus	Ne	I	Ne	I	Ne	I	Ne	I	Ne	I
Camsin M5	1.9600	0.6829	1.2620	0.3622	1.9727	0.6862	1.3349	0.4176	1.3349	0.4176
Camsin M2	1.6374	0.5779	1.2620	0.3622	1.5622	0.5456	1.4098	0.4660	1.7785	0.6295
Camjap A1	1.2523	0.3541	1.9018	0.6671	1.9360	0.6765	1.4235	0.4741	1.5414	0.5360
Camjap A4	1.19801	0.3046	1.1980	0.3046	1.8615	0.6555	1.5414	0.5360	1.6575	0.5860
TM 51	1.2620	0.3622	1.8408	0.6492	1.9931	0.6914	1.8408	0.6492	1.2620	0.3622
TM 134	1.5622	0.5456	1.5622	0.5456	1.8408	0.6492	1.2620	0.3622	1.5622	0.5456
A 37	1.2462	0.3488	1.9756	0.6870	1.8740	0.6592	1.2462	0.3488	1.3376	0.4195
A 47	1.3376	0.4195	1.3376	0.4195	1.7153	0.6077	1.3243	0.4101	1.4322	0.4792
Mean	**1.4320**	**0.4105**	**1.5425**	**0.4997**	**1.9727**	**0.6862**	**1.4429**	**0.4580**	**1.4883**	**0.4970**
St. Dev.	**0.2493**	**0.1270**	**0.3009**	**0.1457**	**1.5622**	**0.5456**	**0.1815**	**0.0923**	**0.1665**	**0.0867**

For each loci, *Ne* = Number of effective alleles, *I* = Shannon's information index

Table 4.7: Analysis of Molecular Variance (AMOVA) for 5 *Camellia* populations based on 8 loci

Source of variation	Df	SS	Estimated variance components	%Variation	Fixation index	P-value
Among populations	4	4.820	0.019	3.00	Fst:0.032	0.001
Among individuals within populations	83	46.902	0.000	0.00	Fis:-0.022	0.692
Within individuals	88	52.000	0.591	97.00	Fit:0.011	0.369
	175	103.722	0.610	100%		
Nm = 6.5624						

Df = degrees of freedom; SS=sum of squares

Table 4.8: Genetic diversity studies of individual families/stocks of interspecific tea hybrids

Families/Stocks																						
Locus	31 (n=7)		73 (n=5)		306 (n=4)		570 (n=2)		597 (n=6)		645 (n=3)		688 (n=12)		691 (n=2)		845 (n=6)		862 (n=12)		921 (n=2)	
	Ne	I	Ne	I	Ne	I	Ne	I	Ne	I	Ne	I	Ne	I	Ne	I	Ne	I	Ne	I	Ne	I
Camsin M5	1.26	0.36	1.26	0.36	1.26	0.36	1.34	0.58	1.64	0.36	1.97	0.69	1.64	0.58	1.34	0.42	1.49	0.51	1.34	0.42	1.64	0.58
Camsin M2	1.26	0.36	1.33	0.42	1.26	0.36	1.41	0.30	1.19	0.36	1.56	0.55	1.19	0.30	1.69	0.60	1.71	0.61	1.34	0.42	1.26	0.36
Camjap A1	1.25	0.35	1.42	0.47	1.90	0.67	1.42	0.59	1.66	0.64	1.94	0.68	1.66	0.57	1.86	0.66	1.48	0.51	1.86	0.66	1.90	0.67
Camjap A4	1.20	0.31	1.54	0.54	1.20	0.31	1.54	0.25	1.15	0.54	1.86	0.66	1.15	0.25	1.88	0.66	1.98	0.69	1.60	0.56	1.71	0.61
TM 51	1.41	0.47	1.71	0.61	1.84	0.65	1.84	0.47	1.41	0.66	1.20	0.70	1.41	0.47	1.56	0.55	1.94	0.68	1.94	0.68	1.56	0.55
TM 134	1.41	0.47	1.85	0.65	1.56	0.55	1.26	0.55	1.56	0.55	1.85	0.65	1.56	0.55	1.71	0.61	1.99	0.69	1.56	0.55	1.56	0.55
A 37	1.11	0.20	1.34	0.42	1.20	0.69	1.25	0.57	1.62	0.46	1.87	0.66	1.62	0.57	1.53	0.53	1.25	0.35	1.62	0.57	1.53	0.53
A 47	1.25	0.35	1.43	0.48	1.34	0.42	1.32	0.42	1.34	0.48	1.72	0.61	1.34	0.42	1.34	0.42	1.62	0.57	1.62	0.57	1.34	0.42
Mean	**1.27**	**0.36**	**1.52**	**0.51**	**1.47**	**0.52**	**1.43**	**0.45**	**1.42**	**0.53**	**1.71**	**0.64**	**1.42**	**0.45**	**1.65**	**0.58**	**1.71**	**0.59**	**1.65**	**0.57**	**1.55**	**0.53**
St. Dev.	**0.10**	**0.09**	**0.18**	**0.08**	**0.28**	**0.15**	**0.19**	**0.09**	**0.1**	**0.18**	**0.08**	**0.28**	**0.15**	**0.19**	**0.10**	**0.09**	**0.18**	**0.08**	**0.28**	**0.15**	**0.19**	**0.10**

NB: For each loci, *Ne* = Number of effective alleles, *I* = Shannon's information index

4.5.3 Genetic Relationships and Population Structure

Gene flow to the interspecific hybrids was characterized using genetic relationship analysis and parentage analysis. Relative genetic contribution of the wild alleles to interspecific hybrids was estimated using genetic population structure analysis.

For genetic relatedness analysis, Jaccard similarity coefficient values were utilized in identifying genetic relationships or main clusters (Figure 4.6). The matrix was derived from the proportion of shared fragments, which indicates the degree of relatedness among the genotypes (Kosman & Leonard, 2005). Estimated similarity ranged from 2.3% between the most dissimilar individuals, TRFK 303/577 (parental genotype) and TRFK 306/2, to 92.9% between closely related wild accessions *C. brevistyla* and *C. sasanqua* (Figure 4.6). Although the range of similarity coefficient was large, the tested accessions were not clearly separated into distinct clusters. Missing data in some genotypes may account for this. Nevertheless, 5 clusters with 4, 10, 57, 2, and 3 individuals were generated (Figure 4.6). Most accessions were grouped into one large cluster (C3) with several nested sub-clusters. Only twelve accessions namely TRFK 91/2, TRFK 691/2, TRFK 73/5, TRFK 31/38, *C. japonica*, TRFK 31/11, BBK BB35, TRFK 688/18, AHP SC12/28, EPK TN14-3, TRFK 31/32, and TRFK 688/1-2007 were ungrouped at about 50% similarity level.

The dendrogram confirmed the close relatedness among most of the accessions. However, some clusters differed from the conventional classification. For example, Cluster 2 (C2) comprised two subgroups, subgroup 1 (TRFK 73/2, TRFK 73/4, and TRFK 73/4) and subgroup 2 (TRFK 73/1, TRFK 301/1, and TRFK 306/2). Further, Cluster 3 (C3) had two sub-clusters, subgroup 1 comprising TRFK 845/2, TRFK 845/4, TRFK 862/1, and TRFK 845/3 while subgroup 2 had TRFK 688/4, TRFK 862/5, TRFK 845/6, TRFK 688/7, TRFK 688/19, and TRFK 845/1, which was expected as they share one parent – TRFK 91/1. The wild-type accessions were also grouped into two clusters, i.e. Cluster l having *C. irrawandiensis* while Cluster 2 had *C. oleifera*, *C. kissi*, *C. brevistyla*, and *C. sasanqua*. *C. japonica* remained ungrouped. Whereas accession 688/1 from the Genet 3c/2005 trial grouped into C3, an accession with the same code-name from Genet 3c/2007 trial remained ungrouped, even though they share parents – TRFK 91/1 x TRFK 303/577.

From the genetic structure analysis, six groups were inferred. Genet 3c/1999 was characterized by the high relative contribution of the 'wild' alleles, wherein 19 individuals (95%) exhibited a clearly predominant 'wild' subpopulation compared to 23 (71.8%) for Genet 3c/2007 and 7 (38.9%) for Genet 3c/2005 (Figure 4.7). In total, the 'wild' genetic configuration was expressed in 49 hybrids, accounting for 70% of total 'wild' genetic contribution. The hybrid population is therefore highly admixed.

4.5.4 Parentage Analysis

Of the 88 tested genotypes, 46 were full sibs from 10 different families (St. 570, St. 597, St. 599, St. 600, St. 660, St. 691, St. 645, St. 862, St. 688, and St. 845). A further 24 were half sibs and seedling selections from 9 families (St. 667, St. 680, St. 921, St. 306, St. 73, St. 31, 91/2, 83/1, and 14/1). Across all simulations using the Cervus 3.0.7 program, low parentage assignment rates were obtained for all scenarios under relaxed confidence levels. No mother was assigned to an offspring at strict confidence levels. Information regarding simulation confidence levels, simulation parameters, log-likelihood (LOD) distributions, and breakdown of parentage assignment for all families is provided in Appendix 4-7.

4.5.4.1 Analysis of Maternity

Parentage analysis correctly identified all the 46 progenies as full sibs though wrongly assigned a maternal parent to 23 of the offsprings and failed to assign any maternal parent to 9 individuals (Table 4.9). Ten of the correct mother-offspring pairs had LOD score of over 0.8 (80% confidence threshold) while four had 0.6-0.7, with a higher value denoting a greater likelihood. Generally, genotypes of known mothers were provided in all analyses. Considering individual families, correct identification of a mother was highest among St. 845 offspring, where three (TRFKs 845/2, 845/4 and 845/6) out of six were assigned a known maternal parent (TRFK 91/1) with LOD score of over 0.7. Four of the 12 offspring from St. 862 family (TRFK 862/5, TRFK 862/4, TRFK 862/3, and TRFK 862/1) were also correctly assigned a mother (TRFK 91/1). The maternal parent (TRFK 301/4) was identified for two individuals (TRFKs 645/14 and 645/5) of the three individuals from St. 645 and two (TRFK 597/15 and TRFK 597/12) of the six offspring from St. 597. Of the two St. 570 offspring, one (TRFK 570/1) was assigned the correct parent. The correct maternal parents, namely TRFK 301/3 and K-purple for TRFK 599/2 and TRFK 660/1, respectively were identified.

Among the 24 half-sibs, 11 had the maternal parent identified under relaxed confidence, 4 had a likely maternal parent without an assigned parentage, and 9 were not assigned any parent (Table 4.10). The LOD score for offspring with an identified maternal parent ranged between 0.43 and 0.78. For TRFK 667/3, two possible mothers were assigned though TRFK 303/577 had higher pair confidence (LOD score) making it the most likely mother than AHP SC12/28 (0.434 vs. 0.399, respectively). Similarly, TRFK 921/1 was assigned three maternal parents (TRFK 303/577, BBK 35, and TRFK K-purple) with TRFK 303/577 having highest LOD score. Nine cultivars comprising five offspring from St. 73 (TRFK 73/1, TRFK 73/2, TRFK 73/3, TRFK 73/4 and TRFK 73/5) were assigned two different mothers: AHP S15/10 and its progeny clone AHP SC12/28. Offspring TRFK 301/1, TRFK 14/1, and TRFK 91/2 were assigned the same candidate maternal parent, AHP SC12/28, with LOD score of 0.785 (Table 4.10). Cultivar TRFK K-purple was incorrectly assigned TRFK 306/4 as the most likely maternal parent under relaxed confidence. However, the other progenies in St. 306 were not assigned maternal parents.

Table 4.9: Predicted candidate maternal parent for full sibs (known mothers were provided to Cervus for this analysis)

Family / Stock	Offspring ID	Crosses	Known mother	Candidate mother ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair confidence
570	570/1	TRFK 301 x *C. japonica*	TRFK 303/1	TRFK 303/1	2	0	7.98E-01	+
	570/2	TRFK 301 x *C. japonica*	TRFK 303/2	TRFK 303/2	0	0	0.00E+00	-
	597/26	TRFK 91/1 x AHP S15/10	TRFK 91/1	TRFK91/1	0	0	0.00E+00	
	597/17	TRFK 91/1 x AHP S15/10	TRFK 91/1	TRFK91/1	1	0	2.83E-01	
	597/15	TRFK 91/1 x AHP S15/10	TRFK 91/1	TRFK91/1	1	0	1.79E-01	+
	597/12	TRFK 91/1 x AHP S15/10	TRFK 91/1	TRFK91/1	2	0	1.79E-01	+
	597/8	TRFK 91/1 x AHP S15/10	TRFK 91/1	TRFK91/1	1	0	1.79E-01	-
	597/1	TRFK 91/1 x AHP S15/10	TRFK 91/1	TRFK91/1	1	0	2.83E-01	
599	599/2	TRFK 91/1 x TRFK 301/3	TRFK 91/1	TRFK 301/3	3	0	1.08E+00	-
600	600/3	TRFK 91/1 x BBK BB35	TRFK91/1	TRFK 91/1	6	0	2.28E+00	-
660	660/1	TRFK K-purple x AHP SC12/28	K-purple	K-purple	2	0	7.98E-01	+
645	645/14	TRFK 301/4x K-purple	TRFK 301/4	TRFK 301/4	3	0	7.21E-01	+
	645/6	TRFK 301/4x K-purple	TRFK 301/4	TRFK 301/4	2	0	4.38E-01	-
	645/5	TRFK 301/4x K-purple	TRFK 301/4	TRFK 301/4	4	0	1.00E+00	-
862	862/5	TRFK 91/1 x TRFK 301/4	TRFK 91/1	TRFK 91/1	4	0	8.36E-01	+
	862/4	TRFK 91/1 x TRFK 301/4	TRFK 91/1	TRFK 91/1	4	0	9.00E-01	+
	862/3	TRFK 91/1 x TRFK 301/4	TRFK 91/1	TRFK 91/1	5	0	1.30E+00	-
	862/1	TRFK 91/1 x TRFK 301/4	TRFK 91/1	TRFK 91/1	3	0	6.17E-01	+
688	688/19	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	1	0	1.16E-01	-
	688/18	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	2	0	2.01E-01	-
	688/13	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	2	0	2.01E-01	-
	688/12	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	2	0	2.01E-01	-
	688/11	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	1	0	8.54E-02	+

	688/10	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	2	0	2.01E-01	-
	688/7	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	1	0	8.54E-02	+
	688/6	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	2	0	2.01E-01	-
	688/4	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	1	0	8.54E-02	+
	688/1-07	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	2	0	2.01E-01	-
	688/15	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	1	1	0.00E+00	
	688/1-05	TRFK 91/1 x TRFK 303/577	TRFK 91/1	TRFK 303/577	1	0	8.54E-02	+
845	845/6	TRFK 301/4 x TRFK 91/1	TRFK 301/4	TRFK 301/4	4	0	1.00E+00	-
	845/5	TRFK 301/4 x TRFK 91/1	TRFK 301/4	TRFK 301/4	0	0	0.00E+00	
	845/4	TRFK 301/4 x TRFK 91/1	TRFK 301/4	TRFK 301/4	3	0	7.21E-01	-
	845/3	TRFK 301/4 x TRFK 91/1	TRFK 301/4	TRFK 301/4	0	0	0.00E+00	
	845/2	TRFK 301/4 x TRFK 91/1	TRFK 301/4	TRFK 301/4	3	0	7.21E-01	-
	845/1	TRFK 301/4 x TRFK 91/1	TRFK 301/4	TRFK 301/4	0	0	0.00E+00	

For each offspring, + represents the most likely mother for relaxed confidence, - is shown for a most likely candidate parent not assigned parentage, and a blank means the candidate parent is not the most likely.

Table 4.10: Predicted candidate maternal parent for half sibs (mothers were not provided to Cervus for this analysis)

Family	Offspring ID	Candidate mother ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair confidence
667	TRFK 667/3	TRFK 303/577	2	0	4.34E-01	+
	TRFK 667/3	AHP SC12/28	1	0	3.99E-01	-
680	TRFK 680/2	TRFK 303/577	1	0	3.99E-01	-
921	TRFK 921/1	TRFK 303/577	2	0	5.78E-01	+
	TRFK 921/1	BBK BB35	1	0	1.79E-01	-
	TRFK 921/1	TRFK K-purple	1	0	1.79E-01	-
301	TRFK 301/1	AHP SC15/10	2	0	7.85E-01	+

14	TRFK 14/1	AHP SC12/28	2	0	7.85E-01	+
91	TRFK 91/2	AHP SC15/10	2	0	7.85E-01	+
73	TRFK 73/5	AHP SC12/28	2	0	7.85E-01	+
	TRFK 73/4	AHP SC12/28	2	0	7.85E-01	+
	TRFK 73/3	AHP SC15/10	2	0	7.85E-01	+
	TRFK 73/2	AHP SC12/28	2	0	7.85E-01	+
	TRFK 73/1	AHP SC15/10	2	0	7.85E-01	+
306	TRFK 306/4	K-purple	2	0	7.85E-01	+

NB: + means the most likely mother for relaxed confidence, - is shown for a most likely candidate parent not assigned parentage.

4.5.4.2 Analysis of Paternity

Results from paternity analysis for full-sibs revealed candidate fathers similar to known fathers for 7 out of 10 families (Table 4.11). It was possible to confirm the father for 12 of the 46 full-sib offspring under strict and relaxed confidence levels. For these clones, pair confidence was generally high, with the LOD score > 0.6. *C. japonica* was identified and assigned as the likely father of TRFK 570/2 but was not assigned to TRFK 570/1 due to low pair confidence (LOD score). For a similar reason, AHP S15/10, which is the known paternal parent for 597 family, was identified as the correct father of all six offspring (TRFK 5971/26, TRFK 597/17, TRFK 597/15, TRFK 597/12, TRFK 597/8, and TRFK 597/1) but was not assigned parentage. In 599 family, TRFK 301/3 was correctly identified as the paternal parent to TRFK 599/2, while BBK BB35 was assigned as the likely father to TRFK 600/3 in the 600 family under strict confidence levels. The known father of the 660 progeny (AHP SC12/28) was correctly assigned to its offspring TRFK 660/1. No paternal parent was assigned to TRFK 691/1 though *C. japonica* was identified as the likely father.

Among St. 645 progenies, only one of the three offspring (TRFK 645/5) was assigned the known father, K-purple, but was identified as the likely paternal parent to the other two – TRFK 645/14 and TRFK 645/6 (Table 4.11). In contrast, the three offspring in 862 family, i.e., TRFK 862/5, TRFK 862/4, TRFK 862/3, and TRFK 862/1, were correctly assigned their known father, TRFK 301/4 (Table 4.11). In contrast, it was not possible to assign a paternal parent to 12 offspring in the 688 family, though TRFK 303/577 was identified as the likely but unassigned father to all offspring except TRFK 688/1-05. Paternity analysis for 845 family yielded mixed results. The known father, TRFK 91/1, was correctly identified and assigned to three of the six offspring – TRFK 845/6, 845/4, and TRFK/2. The other offspring, TRFK 845/5, TRFK 845/3, and TRFK 845/1, had no loci typed and therefore had no had father identified.

For half-sib families, paternity analysis identified a likely father for 15 of the 24 offspring but did not assign any of them parentage (Table 4.12). The likely paternal parent for two clones, TRFK 667/3 and 680/2, from distinct families was identified as TRFK 303/577, with LOD score >0.4. In family 921, TRFK 303/577 was again identified as the likely father of TRFK 921/1 but not TRFK 921/5, which lacked typed loci. Two probable paternal parents were identified for family 73 offspring, i.e., *C. sasanqua* and *C. brevistyla*, but the pair confidence for *C. brevistyla* was higher than that of *C. sasanqua* (LOD = 0.0785 vs. 0.063), making it the most likely father. In family 306, three of the four offspring (TRFK 306/4, TRFK 306/3, and TRFK 306/1) had their paternal parent identified as TRFK 91/1; however, this is the known maternal parent. The identification of the father to clone 306/2 was not achieved as it had no loci typed. Different fathers were identified for other putative hybrid collections. The likely paternal parents to clone TRFK 83/1, TRFK 14/1, and TRFK 91/2 were identified as TRFK 6/8, AHP S15/10, and TRFK 301/4, respectively; though no parentage assignment was achieved.

No maternal parent or paternal parent was identified for three hybrids: TRFK 845/5, TRFK 845/3, and TRFK 845/1. No loci typed were typed for these three hybrids and therefore had no had father or mother identified.

Table 4.11: Predicted candidate fathers for full sibs (known fathers were provided to Cervus for this analysis)

Family	Offspring ID	Crosses	Known father	Candidate father ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair confidence
570	570/1	TRFK 301 x *C. japonica*	*C. japonica*	*C. japonica*	1	0	3.99E-01	-
	570/2	TRFK 301 x *C. japonica*	*C. japonica*	*C. japonica*	2	0	7.98E-01	+
597	597/26	TRFK 91/1 x AHP S15/10	AHP S15/10	AHP S15/10	0	0	0.00E+00	
	597/17	TRFK 91/1 x AHP S15/10	AHP S15/10	AHP S15/10	0	0	0.00E+00	
	597/15	TRFK 91/1 x AHP S15/10	AHP S15/10	AHP S15/10	1	0	1.79E-01	-
	597/12	TRFK 91/1 x AHP S15/10	AHP S15/10	AHP S15/10	1	0	1.79E-01	-
	597/8	TRFK 91/1 x AHP S15/10	AHP S15/10	AHP S15/10	1	0	1.79E-01	-
	597/1	TRFK 91/1 x AHP S15/10	AHP S15/10	AHP S15/10	0	0	0.00E+00	
599	599/2	TRFK 91/1 x TRFK 301/3	TRFK 301/3	TRFK 301/3	3	0	1.08E+00	+
600	600/3	TRFK 91/1 x BBK BB35	BBK BB35	BBK BB35	6	6	2.28E+00	*
660	660/1	TRFK K-purple x AHP SC12/28	AHP SC12/28	AHP SC12/28	2	2	7.98E-01	+
691	691/1	GW Ejulu x *C. japonica*	*C. japonica*	*C. japonica*	1	0	3.99E-01	-
645	645/14	TRFK 301/4x K-purple	K-purple	K-purple	3	0	7.21E-01	-
	645/6	TRFK 301/4x K-purple	K-purple	K-purple	2	0	4.38E-01	-
	645/5	TRFK 301/4x K-purple	K-purple	K-purple	4	0	1.00E+00	+
862	862/5	TRFK 91/1 x TRFK 301/4	TRFK 301/4	TRFK 301/4	4	0	8.36E-01	+
	862/4	TRFK 91/1 x TRFK 301/4	TRFK 301/4	TRFK 301/4	4	0	9.00E-01	+
	862/3	TRFK 91/1 x TRFK 301/4	TRFK 301/4	TRFK 301/4	5	0	1.30E+00	+
	862/1	TRFK 91/1 x TRFK 301/4	TRFK 301/4	TRFK 301/4	3	0	6.17E-01	+
688	688/19	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	1	0	1.16E-01	-
	688/18	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	2	0	2.01E-01	-

	688/13	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	2	0	2.01E-01	-
	688/12	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	2	0	2.01E-01	-
	688/11	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	1	0	8.54E-02	-
	688/10	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	2	0	2.01E-01	-
	688/7	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	1	0	8.54E-02	-
	688/6	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	2	0	2.01E-01	-
	688/4	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	1	0	8.54E-02	-
	688/1-07	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	2	0	2.01E-01	-
	688/15	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	0	0	0.00E+00	
	688/1-05	TRFK 91/1 x TRFK 303/577	TRFK 303/577	TRFK 303/577	1	0	8.54E-02	-
845	845/6	TRFK 301/4 x TRFK 91/1	TRFK 91/1	TRFK 91/1	4	0	1.00E+00	+
	845/5	TRFK 301/4 x TRFK 91/1	TRFK 91/1	TRFK 91/1	0	0	0.00E+00	
	845/4	TRFK 301/4 x TRFK 91/1	TRFK 91/1	TRFK 91/1	3	0	7.21E-01	+
	845/3	TRFK 301/4 x TRFK 91/1	TRFK 91/1	TRFK 91/1	0	0	0.00E+00	
	845/2	TRFK 301/4 x TRFK 91/1	TRFK 91/1	TRFK 91/1	3	0	7.21E-01	+
	845/1	TRFK 301/4 x TRFK 91/1	TRFK 91/1	TRFK 91/1	0	0	0.00E+00	

NB: * represents the most likely mother for strict confidence, + for most likely mother for relaxed confidence, - is shown for a most likely candidate parent not assigned parentage, and a blank means the candidate parent is not the most likely.

Table 4.12: Predicted candidate fathers for half sibs (no fathers were provided to Cervus for this analysis)

Family	Offspring ID	Candidate father ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair confidence
667	TRFK 667/3	TRFK 303/577	2	0	4.34E-01	-
680	TRFK 680/2	TRFK 303/577	1	0	3.99E-01	-
921	TRFK 921/5	0	0	0	0.00E+00	
921	TRFK 921/1	TRFK 303/577	2	0	5.78E-01	-
	TRFK 301/1	AHP SC12/28	1	0	7.85E-02	-
	TRFK 14/1	AHP S15/10	1	0	7.85E-02	-
	TRFK 91/2	TRFK 301/4	2	0	9.85E-02	-
73	TRFK 73/5	*C. sasanqua*	1	0	6.30E-02	-
73	TRFK 73/4	*C. sasanqua*	1	0	6.30E-02	-
73	TRFK 73/3	*C. brevistyla*	1	0	7.85E-02	-
73	TRFK 73/2	*C. brevistyla*	1	0	7.85E-02	-
73	TRFK 73/1	*C. sasanqua*	1	0	6.30E-02	-
306	TRFK 306/4	TRFK 91/1	3	1	5.30E-01	-
	TRFK 306/3	TRFK 91/1	3	1	5.30E-01	-
	TRFK 306/2	0	0	0	0.00E+00	
	TRFK 306/1	TRFK 91/1	3	1	5.30E-01	-
	TRFK 83/1	TRFK 6/8.	2	0	4.58E-01	-

NB: - means most likely candidate parent not assigned parentage, while blank means the candidate parent is not the most likely

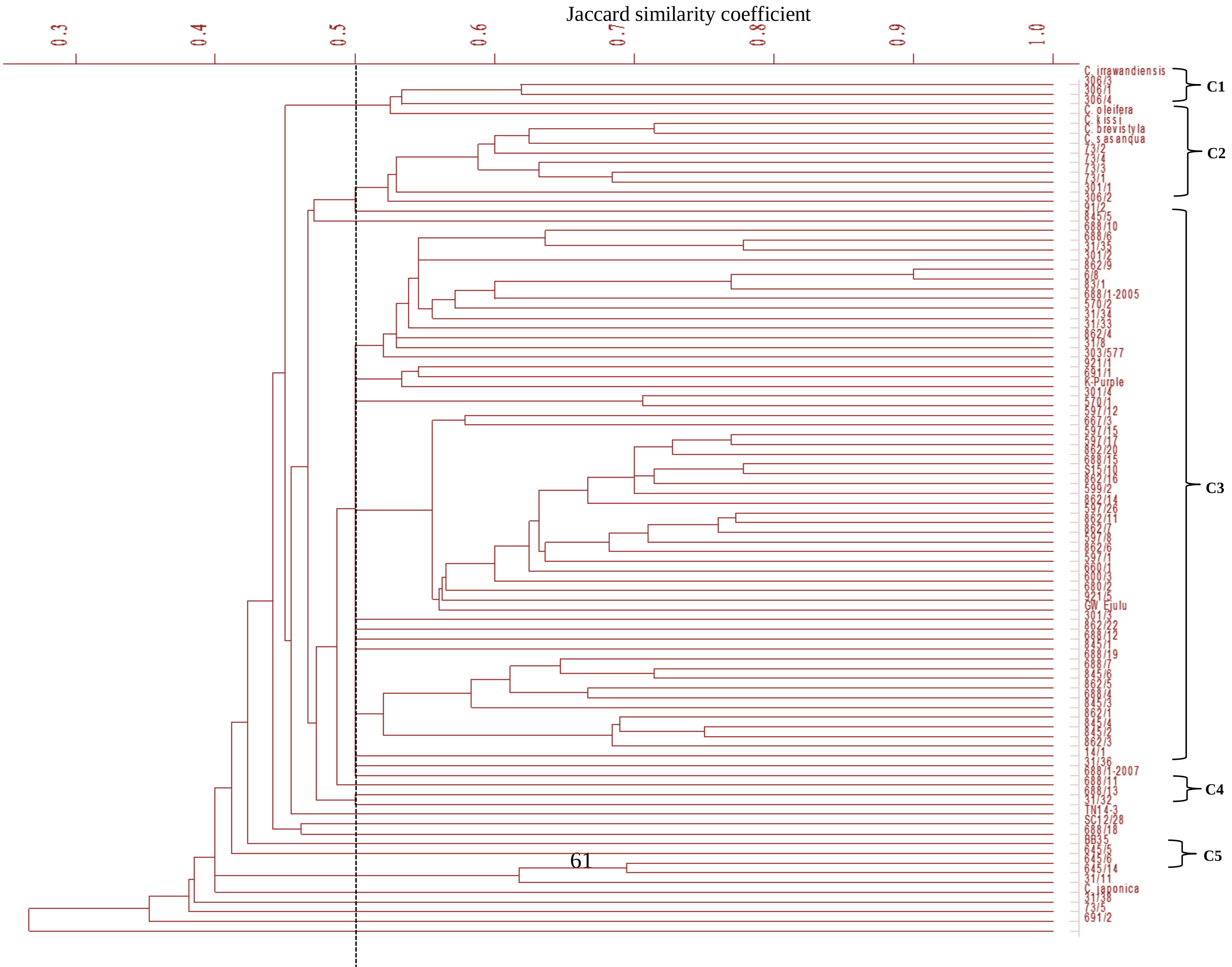

Jaccard similarity coefficient
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
C. irrawandiensis
306/3
306/1
306/4
C. oleifera
C. kissi
C. brevistyla
C. sasanqua
73/2
73/4
73/3
73/1
301/1
306/2
91/2
845/5
688/10
688/6
31/35
301/2
862/9
6/8
83/1
688/1-2005
570/2
31/34
31/33
862/4
31/8
303/577
921/1
691/1
K-Purple
301/4
570/1
597/12
667/3
597/15
597/17
862/20
688/15
S15/10
862/16
599/2
862/14
597/26
862/11
862/7
597/8
862/6
597/1
660/1
600/3
680/2
921/5
GW Ejulu
301/3
862/22
688/12
845/1
688/19
688/7
845/6
862/5
688/4
845/3
862/1
845/4
845/2
862/3
14/1
31/36
688/1-2007
688/11
688/13
31/32
TN14-3
SC12/28
688/18
BB35
645/5
645/6
645/14
31/11
C. japonica
31/38
73/5
691/2
C1
C2
C3
C4
C5

Figure 4.8: Dendrogram illustrating genetic relationships among the 88 accessions of Genet 3c/1999, Genet 3c/2005, Genet 3c/2007, wild, and parental teas generated by the neighbor-joining cluster analysis computed from 8 SSR markers. Paired Jaccard's similarity coefficient was obtained between cultivars and used to construct the dendrogram. Five clusters were obtained at 50% similarity.

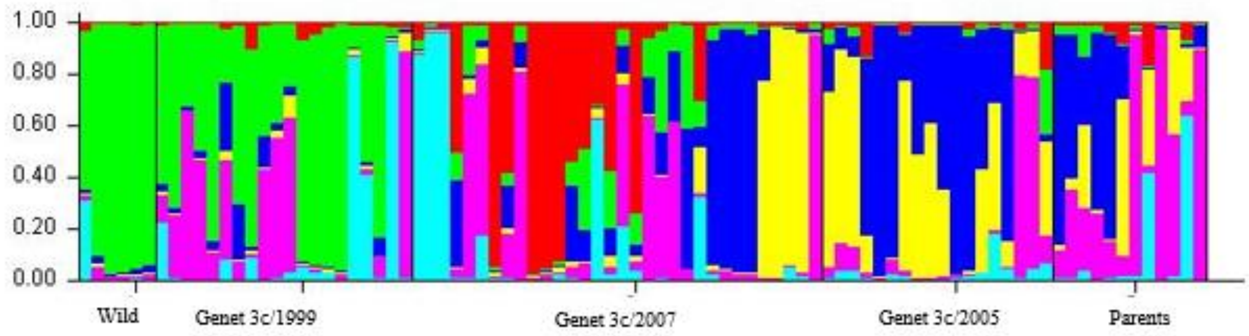


Figure 4.9: A bar plot of population structure analysis results – inferred population structure generated by Structure v. 2.3.4 software according to K = 6 based on eight SSRs. Each vertical bar represents the genome of each individual. Six clusters are inferred: Cluster II (green), Cluster II (pink), Cluster III (turquoise blue), Cluster IV (blue), Cluster V (red), and Cluster VI (yellow).

CHAPTER FIVE

DISCUSSION

5.1 Use of EST-SSR markers in Identification of Interspecific Hybrids of Tea

Two novel EST-SSR markers (Camjap A1 and Camjap A4) exhibited a high polymorphic information content and discriminating power as genomic microsatellites to identify interspecific hybrids of tea. Further, based on EST data, the *Camellia* genome contains a high di-nucleotide repeat density relative to other repeats. Of all di-nucleotide repeat types, (TA)n, (AT)n and (AG)n are the most abundant SSRs in the *Camellia* genome.

Genomic SSR markers have been used extensively to detect genetic variation in tea populations and estimate their genetic diversity (Freeman *et al*., 2004; Ma *et al*., 2010; Yao *et al*., 2012; Wambulwa *et al*., 2016; Dubey *et al*., 2020). Microsatellites or SSR markers are powerful tools for assessment of genetic diversity, gene flow rate and molecular breeding in crops compared to RFLP, RAPD, or AFLP markers due to their multi-allelic nature, codominant inheritance, reproducibility, high variability, and wide genome coverage (Gupta *et al*., 2005; Taheri *et al*., 2018). An increase in sequencing projects has provided a wealth of DNA sequence information that is useful for mining EST-SSR markers for genetic improvement.

Yao *et al*. (2012) developed and utilized 96 polymorphic EST-SSR markers for population structure analysis in 450 Chinese tea accessions. Ma *et al*. (2010) also report the development and polymorphism validation of 74 EST-based SSR markers in 45 tea cultivars belonging to 7 different varieties. In a recent study, 82 SSRs were developed from sequences available in public databases such as ESTs, Genome Survey Sequence (GSS) and RNA-seq, and were validated using 36 tea genotypes (Dubey *et al*., 2020).

In the present study, frequency analysis revealed that di-nucleotide repeats were the most frequent motif type (62.9%) in the wild *Camellia* genomes followed by the tri-nucleotides (Figure 4.1). A high di-nucleotide repeat density (over 50%) relative to the other repeats has been reported in *C. sinensis* ESTs (Sharma *et al*., 2009; Wu *et al*., 2013). The different values used to detect SSR motifs in EST data could explain the variation in the number of reported SSR classes between studies (Dubey *et al*., 2020). The most abundant di-nucleotide repeats were (TA)n, (AT)n and (AG)n, contributing 43.9% of all di-repeats. This confirms reports by Tan *et al*. (2013) that the AG/CT motif is the most frequent repeat unit followed by AT/TA in *C. sinensis*. Higher DNA polymerase-mediated slippage events in shorter units can explain this variation in microsatellite density (Kruglyak *et al*., 2000). *PIC* values that estimate the informativeness of a marker based on the allelic frequency and total alleles detected (Nagy *et al*., 2012; Reyes-Valdés, 2013) were not significantly correlated with allele frequency data (r = 0.49, p = 0.28), suggesting that the usefulness of a marker was not dependent on detecting a higher number of alleles. When the *PIC* value exceeds 0.5, it indicates informativeness (Botstein *et al*., 1980). Three SSR markers comprising of one novel EST-SSR marker (Camjap A4) and two adapted markers (A37 and A47) had an average value of 0.50, which is considered informative in studying genetic diversity in inter- and intra-specific hybrids (Table 4.3). Generally, the average *PIC* value for genomic microsatellites (core markers of *C. sinensis*) was relatively higher (0.275) than that of the novel EST-SSR markers (0.250) though not significantly different ($p \leq 0.05$). Genomic SSRs exhibit high polymorphism levels and occur widely in the genome but are less transferable between species (Kuleung *et al*., 2004; Parthiban *et al*., 2018). In contrast, EST-SSRs are less polymorphic (Decroocq *et al*., 2003) than the genomic SSRs because they occur in the transcribed region that is highly conserved (Cho *et al*., 2000).

Discriminating power (D) is also a useful estimator of the informativeness of a marker (Amiryousefi *et al*., 2018). SSR markers with higher discriminating power ($D \geq 0.7$) give an optimal primer combination for discriminating cultivars (Tessier *et al*., 1999). In this study, the D values of two primers, namely Camjap A1 and TM 134, were 0.70 and 0.75, respectively (Table 4.3). These SSR markers are thus efficient tools for definitive identification of inter- and intra-specific hybrids. Both polymorphic information content and discriminating power are dependent on allele frequency. However, the discriminating efficiency of a primer is not exclusively dependent on the number of polymorphic bands it produces. This implies that SSRs with similar polymorphic patterns can have different discriminating powers, e.g., TM 134 and A37 (Table 4.3). On the other hand, two markers producing significantly different numbers of polymorphic bands may have fairly similar discriminatory powers, e.g., Camjap A4 and Camsin M2. Frequency differences in banding patterns produced with these primers could explain this result (Tessier *et al*., 1999).

Recent reports have suggested that SSR markers are efficient tools for studying diversity in closely related breeding lines (Zhang *et al*., 2018; Zhou *et al.*, 2019). Wambulwa *et al*. (2016) used 23 polymorphic SSR loci to separate East African teas into groups based on geographical origins. In these studies, *PIC* and related indices of polymorphism were used as a benchmark for assessing the effectiveness of SSR markers. Our published data indicate that the EST-SSR markers (Camjap A1 and Camjap A4) and genomic microsatellites (TM 58, TM 134, A37, A47, Camsin M2, and Camsin M5) tested can effectively be used to discriminate interspecific hybrids of tea as well as identify germplasm to include in tea improvement programs (doi:10.20425/ijts1515).

5.2 Estimation of Genetic Diversity of Interspecific Tea Hybrids

The genetic diversity in the interspecific tea hybrids was low to moderate, with Genet 3c/2007 population exhibiting the highest genetic diversity (I = 0.6862). Further, full-sibs also showed a higher genetic diversity than half-sibs. The level of variation between populations was low suggesting close relationships among the hybrids. Relationship analysis revealed five major clusters, with 59 of the 88 cultivars grouping in one cluster along with three wild type species.

Previous studies have shown that over-reliance on a few breeding stocks reduces genetic diversity in cultivated germplasm (Wachira *et al*., 2001; Chen *et al.*, 2004; Yao *et al*., 2012). Thus, molecular characterization of the existing gene pool is required to identify disparate genotypes for inclusion in breeding programs while eliminating duplicates. In this study, SSR marker analysis revealed significant genetic diversity across the eight loci analyzed by Shannon's diversity index. The index estimates genetic diversity within and among subpopulations and varies between 0 and 1, with values closer to zero, indicating lower genetic diversity (NIST, 2016). Additionally, the number of effective alleles (*Ne*) gave expected heterozygosity or gene diversity at each locus (Nei, 1973). Genetic diversity was highest at TM 51 locus ($I = 0.5838$, $Ne = 1.6522$) and lowest at Camsin M5 ($I = 0.4524$, $Ne = 1.3876$) (Table 4.5). The mean values ($I = 0.5216$, $Ne = 1.5185$) indicated that genetic distances among the clones were larger with wide genetic base. These results were consistent with Liu *et al*. (2012) who reported average Shannon information index of 0.5586 and Nei's genetic diversity of 0.3797 in wild tea accessions.

Among the five populations studies, genetic diversity was moderate ($I = 0.5216$), with Genet 3c/2007 population being the most genetically diverse population ($I = 0.6862$) and wild tea accessions the least diverse ($I = 0.4105$) (Table 4.5), possibly due to lower genetic perturbations in wild teas (Niu *et al*., 2019). Further, low genetic diversity in the wild populations indicates stronger effects of genetic drift due to domestication for breeding purposes (Zhao *et al*., 2014). The new alleles incorporated through interspecific crosses are expected to increase genetic diversity in the progenies compared to wild tea accessions.

Alternatively, high genetic diversity among the interspecific hybrids could be linked to the biological characteristics of tea. As the plant is highly self-incompatible, natural outbreeding with wild relatives increase genetic variability (Ellstrand *et al*., 1999). Full-sibs were also more genetically diverse than half-sibs (($Ne = 1.5675$, $I = 0.5425$ vs. ($Ne = 1.4200$, $I = 0.4633$). This suggests that controlled bi-parental mating involving disparate breeding stocks creates more genetic variability in tea than open pollination.

Genetic variation among the subpopulations also varied significantly across the eight loci but the overall population differentiation was moderate ($F_{ST} = 0.0661$). Camsin M5 (Fst = 0.1656) being the most differentiated locus, while TM 134 (Fst = 0.0115) was the least differentiated locus. The F_{ST} among the five populations was 0.032 (Table 4.7), which is lower than a value of 0.101 obtained in 415 accessions from four population groups comprising of pure wild type, admixed wild type, ancient landraces, and modern landraces from Guizhou, China (Niu *et al*., 2019). The difference is attributed to lower genetic exchanges among the isolated natural Chinese populations studied, causing high genetic differentiation, compared to the less differentiated hybrid populations that share parental lines (Yao *et al.*, 2012).

Gene flow involves the transfer of alleles between two populations of a species, and therefore, it is a useful tool for analyzing population processes within and between species (Gerber *et al*., 2014). The introduction of new alleles in a population where none existed previously is an important source of variation (Futuyma, 1998). High gene flow observed in the present study can also accounts for the moderate genetic differentiation. Zong *et al*. (2015) considered that Nm >1 may indicate the occurrence of gene exchange. The gene flow averaged (6.5624 with SD = 6.3670) at the eight loci, suggests extensive genetic exchanges among the populations studied. The high gene flow can be attributed to self-incompatibility mating system in tea (Ellstrand *et al*., 1999).

AMOVA revealed a higher distribution of genetic variation (97%) within individual genotypes than among populations (3%) (Table 4.7). Similar studies by Chen *et al*. (2005) reported lower variation (4.6%) among different taxa based on allozyme markers. While Wachira *et al*. (2001) reported 72% variation in individuals within populations of *C. sinensis* and wild *Camellia* species based on AFLP and RAPDs markers, Kaundun and Park (2002) reported 16% diversity among populations of Korean tea using RAPD markers. Higher average gene flow rate (Nm = 6.5624) among the populations might have reduced their genetic differentiation.

The high gene flow may be attributed to the outcrossing nature and the self-incompatible mating system of tea (Ellstrand *et al*., 1999). Further, allogamy and high outcrossing rates promote the maintenance of high within-population diversity, while hindering genetic variability among populations in alfafa genotypes (Rhouma *et al*., 2014). Interestingly, intra-population variability was lacking among the hybrids, indicating that the genetic diversity is preserved within individuals. Research has suggested high gene flow, natural selection, and the breeding system as the main evolutionary factors affecting genetic variation within populations (Hamrick *et al*., 1992; Zhao *et al*., 2014). In this study, as a consequence of intraspecific breeding and high gene flow between wild *Camellia* species and cultivated tea, the hybrid cultivars were composed of many different genotypes, but variation between the populations was low. Since four (TRFK 303/577, TRFK 301/4, TRFK 91/1, and TRFK K-purple) out of the nine maternal parents (44.4%) used were common across the crosses, such low inter-population variation was expected.

The genetic relatedness of the *Camellia* individuals using the Neighbor-Joining analysis was consistent with the population subdivisions identified using STRUCTURE analysis (Figure 4.7). However, 12 accessions, i.e., TRFK 91/2, TRFK 691/2, TRFK 73/5, TRFK 31/38, *C. japonica*, TRFK 31/11, BBK BB35, TRFK 688/18, AHP SC12/28, EPK TN14-3, TRFK 31/32 and TRFK 688/1-2007, were detached from the five clusters identified at about 50% similarity. Based on individual families, three St. 306 clones (TRFK 306/1, TRFK 306/3 and TRFK 306/4) grouped in Cluster 1 separately from TRFK 306/2 that grouped in Cluster 2. Similarly, TRFK 73/5 did not cluster with TRFK 73/1, TRFK 73/2, TRFK 73/3, and TRFK 73/4 in Cluster 2. Stutter products visualized as variable band lengths produced due to SSRs replication slippage during *in vitro* amplification may account for this difference (Hosseinzadeh-Colagar *et al*., 2016).

Among wild-type individuals, four species i.e. *C. oleifera, C. kissi, C. sasanqua,* and *C. brevistyla* grouped together in Cluster 2, while *C. irrawandiensis* grouped in Cluster 1, and *C. japonica* was ungrouped. Consistent with these results, *C. kissi, C. brevistyla*, and *C. sasanqua* have been shown to be closely related using RAPD markers (Wachira *et al*., 1997). In another study, *C. brevistyla* groups with *C. kissi*, and *C. oleifera* (Su *et al*., 2017). As both *C. irrawandiensis* and *C. japonica* were grouped separate, they could be genetically distant from the other wild individuals.

As expected, most half- and full-sibs grouped in Cluster 3 (Figure 4.6) along with their parents (44.4% of parents are shared). In total, 21 individuals from Genet 3c/2007, 10 from Genet 3c/2005, 8 from Genet 3c/1999, and 7 parents grouped in cluster 3, indicating common pedigree. Cultivars TRFK 688/7, TRFK 688/19, TRFK 845/1, and TRFK 688/12 joined the cluster rather early at about 0.5 or 50% similarity level, whereas TRFK 301/2 and TRFK 862/9 joined cluster 3 later than any of the cultivars at 0.9 or 90% similarity. Cultivars TRFK 688/11 and TRFK 688/13 grouped together with TRFK 31/32 in Cluster 4. Closely grouped in cluster 5 were three progenies of St 645 i.e. TRFK 645/4, TRFK 645/6 and TRFK 645/4. Four hybrids– TRFK 691/2, TRFK 73/5, TRFK 31/38, and TRFK 31/11 – were the most genetically distant with a similarity coefficient of less than 0.4.

5.3 Genetic Contribution of Wild Tea Species to Cultivated Tea Germplasm

The wild allele configuration varied between hybrid populations and was highest in Genet 3c/1999 population and lowest in Genet 3c/2005. The interspecific hybrids were only moderately genetically differentiated and a high gene flow was detected among subpopulations. In addition, similar maternal and paternal parents were identified for both half-sibs and full sibs suggesting shared parentage.

Species of *Camellia* have been shown to readily hybridize among themselves, indicating a close relationship typical of ecospecies (Wachira *et al*., 1997). In structure analysis, most intraspecific hybrids in Genet 3c/1999 exhibited a high contribution of the wild type to their genetic constitution (Figure 4.7). In total, 95% of the progenies in this trial share their wild genetic makeup compared to 71.8% for Genet 3c/2007 and 38.9% for Genet 3c/2005. Overall, wild genotypes contributed 70% of alleles in the sampled population, suggesting introgressive hybridization into the cultivated gene pool (Wachira *et al*., 1997). The high similarity between *C. irrawandiensis* and St. 306 hybrids (306/3, 306/1 and 306/4) confirms that these clones were hybrids between *C. irrawandiensis* and *C. sinensis* var. *sinensis*. Both the parent and progenies are rich in anthocyanin pigments, making the leaves appear purple (Wachira *et al*., 1997)

A high gene flow rate (6.5624 individuals per generation) among the studied genotypes was observed (Table 4.7). Breeding a pure line may be achieved as the high long-term gene flow produces a less genetically differentiated population (Zhang et al., 2019). Gene flow increases genetic uniformity in hybrids bred from similar parental lines. However, true breeding to propagate pure lines is not attainable as tea is an outcrossing heterozygous plant (Hazra et al., 2018). *C. irrawandiensis, C. oleifera, C. kissi, C. brevistyla*, and *C. sasanqua* clustered with interspecific hybrids. However, most hybrids of Genet 3c/2007 and Genet 3c/2005 trials showed little association with the six wild type accessions, indicating limited admixture events among these taxa. Of the six wild ecospecies, *C. japonica* had no genetic contribution to the cultivated gene pool (Figure 4.7).

Parentage analysis revealed the likely maternal and paternal parents to half- and full sib clones. Among full-sib families, the known mothers were correctly assigned to all members of stocks 845, 599, and 660, two of three clones from St. 645, two of the six clones from St. 597, and one of the two clones from St. 570 (Table 4.9). Allele size differences due to replication slippage during PCR may explain the unexpected patterns in paternity for clones not assigned the correct mother (Hosseinzadeh-Colagar *et al*., 2016). The correct paternity was confirmed for 12 of the 46 full-sib offspring (Table 4.11). Paternal parents were correctly identified for the remaining clones although they were not assigned due to low paired LOD score (Kalinowski *et al*., 2007). Of the 24 half-sib offspring, 11 cultivars were correctly assigned their maternal parents, while a likely mother was identified but not assigned to 4 clones (Table 4.10). A maternal parent was neither identified nor assigned to 9 half-sib progenies. Interestingly, two candidate mothers, i.e., TRFK 303/577 and AHP SC12/28, were identified for TRFK 667/3, a progeny of Taiwan Yamacha 87 (not included in the study), implying potential ancestral admixture with *var. assamica* (Yamashita *et al*., 2019). This would also account for the multiple candidate mothers (TRFK 303/577, BBK 35, and TRFK K-purple) assigned to three offspring in St. 921, whose known maternal parent is TRFK 91/1. Two likely mothers half-sibs St. 73 were identified as AHP SC15/10 and its progeny clone AHP SC12/28 (Wachira & Kamunya, 2017). Cultivar TRFK K-purple was identified, though incorrectly, as the potential candidate maternal parent to TRFK 306/4 – the known maternal parent is TRFK 91/1, implying a common pedigree with St. 306. The paired

LOD score of representative clones from St. 306, i.e., TRFK 306/1, TRFK 306/2, and TRFK 306/3 was not adequate for parentage assignment, probably due to fewer typed loci (Kalinowski *et al*., 2007).

Fewer (12) full-sib progenies had their known paternal parents identified than half-sib offspring which could be attributed to the comparatively lower typed loci matching those of known fathers (Kalinowski *et al*., 2007). The pair confidence was high for clones with confirmed paternity, LOD > 0.6. In contrast, the father to offspring of most half-sib families were identified but were not assigned due to low paired LOD score (Table 4.12). Notably, *C. sasanqua* and *C. brevistyla* were identified as paternal parents of St. 73. The identification of alternative paternal parents, for example in St. 73, and a common paternal parent like cultivar TRFK 303/577 for different families such as TRFK 667/3 and TRFK 680/2 (Table 4.12), suggests that the paternal parents analyzed are closely related. Also, missing data at some male loci could lead to parentage assignment errors. Generally, the number of loci for both paternity and maternity analysis was low (≤ 6), hence it is unlikely to account for all alleles inherited from either the mother or father.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

6.1.1 Use of EST-SSR Markers in Identification of Interspecific Hybrids of Tea

The eight SSR loci (two novel EST-SSRs and six adapted microsatellites) verified in the present study could be useful polymorphic markers in characterizing tea hybrids. They were used to efficiently genotype 70 full- and half-sib progenies alongside 12 maternal parents and 6 wild-type accessions, giving novel insights into their genetic diversity and population structure. The eight SSRs were selected from a set of 20 which exhibited mean *PIC* of 0.40 and discriminating power of 0.30 in the initial screening with three interspecific cultivar and one intraspecific cultivar. There were no significant difference in the average PIC values of adapted genomic microsatellites (0.275) and novel EST-SSRs (0.250). The two novel loci developed from ESTs were detected in hybrid and parental clones as well as in the six wild *Camellia* species, indicating their cross-species transferability and potential use in marker-assisted selection.

6.1.2 Estimation of Genetic Diversity of Interspecific Tea Hybrids

Genetic diversity of interspecific hybrids based on SSR markers varied between populations. Overall, the hybrid populations were found to be more genetically diverse than the wild tea population due to genetic admixture situation during breeding. The low-to-moderate genetic diversity in the families studied suggests shared or closely related paternal parents. Most variation was found within individuals than among the population, while the entire population was only moderately differentiated (F_{ST} = 0.0661), which is attributed to a high genetic introgression among the five populations. Most of the accessions grouped together into expected clusters (based on conventional classification), except *C. irrawandiensis* and 306/2 that grouped in separate clades from their known lineage, while *C. japonica* and nine other hybrids remained ungrouped even at 50% similarity. This suggests a wide genetic base of individual hybrid families, wild-type species, and parental accessions.

6.1.3 Genetic Contribution of Wild Tea Species to Cultivated Tea Germplasm

Although the parentage analysis suggested the possibility of multiple paternities of some clones, this result had low statistical support, and could have resulted from genotyping errors. In population structure analysis, the relative genetic contribution of wild teas in cultivated germplasm differed between the three hybrid populations. The highest wild genetic configuration with 95% of the accessions exhibiting clear wild alleles was in trial Genet 3c/1999, which are half-sib progenies. Paternity analysis demonstrated that extensive genetic exchange occurred between wild tea and cultivated teas, implicating *C. irrawandiensis, C. kissi, C. brevistyla, C. oleifera*, and *C. sasanqua* as putative paternal parents of these progenies. Ungrouping of *C. japonica* indicated that the species has not been extensively used in interspecific hybridization.

6.2 Recommendations

1. Since the EST-SSRs exhibited effective selection of interspecific hybrids similar as genomic microsatellites, the number of EST-SSR loci should be increased for an accurate assessment of genetic diversity.
2. Trait-associated fragments should be sequenced to determine the genes of interest.
3. Four interspecific hybrids namely TRFK 691/2, TRFK 73/5, TRFK 31/38, and TRFK 31/11 were the most genetically distant with a Jaccard similarity coefficient of <0.4. These should be exploited as highly conservation resource for enhanced genetic diversity to reverse the existing genetic bottlenecks resulting from overreliance on a few elite breeding stocks in tea over time.
4. Inclusion of *C. japonica* in future improvement programs would widen the tea genetic scope. Phytochemical characterization has shown that C. japonica contains

several bioactive molecules such as phenolic compounds, terpenoids, and fatty acids that could be introduced into tea through interspecific hybridization (Pereira et al., 2022).

5. Although parentage analysis suggested multiple or shared paternities for half- and full-sib progenies, it failed to correctly assign known parents to the offspring. It is recommended that future research uses a larger number of loci for better precision.

REFERENCES

Amiryousefi, A., Hyvönen, J., & Poczai P. (2018). IMEC: Online marker efficiency calculator. *Applications in Plant Science*, *6*(6), 1-4. https://doi.org/10.1002/aps3.1159

Antonova, T., Guchetl, S., Tchelustnikova, T., & Ramasanova, S. (2006). Development of marker system for identification and certification of rice lines and hybrids on the basis of SSR analysis. *Helia, 29*(45), 63-72. https://doi.org/10.2298/HEL0645063A

Balasaravanan, T., Pius, P.K., Kumar, R. R., Muraleedharan, N., & Shasany, A.K. (2003). Genetic diversity among south Indian tea germplasm (*Camellia sinensis, C. assamica* and *C. assamica*spp. *lasiocalyx*) using AFLP markers. *Plant Science, 165*(2), 365-372. https://doi.org/10.1016/S0168-9452(03)00196-1

Bandyopadhyay, T. (2011). Molecular marker technology in genetic improvement of tea. *International Journal of Plant Breeding and Genetics, 5*(1), 23-33. https://dx.doi.org/10.3923/ijpbg.2011.23.33

Banerjee, B. (1992). Botanical classification of tea. In K. C. Willson & M. N. Clifford (Eds.), *Tea: Cultivation to Consumption* (pp. 25-51). Springer.

Barchetia, S., Das, C., Handique, P., & Das, S. (2009). High multiplication frequency and genetic stability for commercialization of the three varieties of micropropagated tea plants (*Camellia* spp.). *Science and Horticulture, 120,* 544-550. https://doi.org/10.1016/j.scienta.2008.12.007

Bekele, A., & Bekele, E. (2014). Overview: Morphological and molecular markers role in crop improvement programs. *International Journal of Current Research in Life Sciences, 3*(3), 035-042.

Biswas, K. P. (2006). *Description of tea plant* In Encyclopedia of medicinal plants. New Dehli: Dominant Publishers and Distributors.

Botstein, D., White, R. L., Skalnick, M. H., & Davies, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics, 32*(3), 314-331. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1686077/pdf/ajhg00189-0020.pdf

Bramel, P. J., & Chen, L. (2019). *A global strategy for the conservation and use of tea genetic resources*. Crop Trust. Retrieved from https://cdn.croptrust.org/wp/wp-content/uploads/2019/04/Global_Strategy_Tea-1.pdf

Chahal, G., & Gosal, S.S. (2002). *Principles and procedures of plant breeding*. Boca Raton, FL: CRC Press.

Chang, H. T., & Bartholomew, B. (1984). *Camellias*. England: Batsford.

Chen, J., Wang, P., Xia, Y., Xu, M., & Pei, S. (2005). Genetic diversity and differentiation of *Camellia* sinensis L. (cultivated tea) and its wild relatives in Yunnan province of China, revealed by morphology, biochemistry and allozyme studies. *Genetic Resources and Crop Evolution, 52*(1), 41-52. https://doi.org/10.1007/s10722-005-0285-1

Chen, X., Hao, S., Wang, L., Fang, W., Wang, Y., & Li, X. (2012). Late-acting self-incompatibility in tea plant (*Camellia sinensis*). *Biologia, 67*(2), 347-351. https://doi.org/10.2478/s11756-012-0018-9

Chen, L., Yang, Y., & Yu, F. (2004). Genetic diversity, relationship and molecular discrimination of elite tea germplasms [*Camellia* sinensis (L.) O. Kuntze] revealed by RAPD markers. *Molecular Plant Breeding,* 2(3), 385-90.

Cho, Y. G., Ishii, T., Temnykh, S., Chen, X., Lopovich, L., McCouch, S. R., Park, W. D., Ayres, N., & Cartinhour, S. (2000). Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics, 100*(5), 713-722. https://doi.org/10.1007/s001220051343

Das, A., & Mishra, R. R. (2020). Economics related to tea (*Camellia sinensis*) sector – An overview. *Agriallis, 2*(12), 18-23.

Decroocq, V., Fave, M. G., Hagen, L., Bordenave, L., & Decroocq, S. (2003). Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theoretical and Applied Genetics, 106*(5), 912-922. https://doi.org/10.1007/s00122-002-1158-z

Devarumath, R. M., Nandy, S., Rani, V., Marimuthu, S., Muraleedharan, N., & Raina, S. N. (2002). RAPD, ISSR and RFLP fingerprints as useful markers to evaluate genetic integrity of micro-propagated plants of three diploid and triploid elite tea clones representing *Camellia sinensis*(china type) and *C. assamicas* sp. *Assamica* (Assam-India type). *Plant Cell Reports, 21*(2), 166-173. https://doi.org/10.1007/s00299-002-0496-2

Dodds, P. N. & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics, 11*, 539–548. https://doi.org/10.1038/nrg2812

Donald, R. (2001). *The science of plant morphology: Definition, history, and role in modern biology*. Berkeley, CA: University of California.

Dubey, H., Rawal, H. C., Rohilla, M., Lama, U., Kumar, P. M., Bandyopadhyay, T., Gogoi, M.,Singh, N. K., Mondal, T. K. (2020). TeaMiD: A comprehensive database of simple sequence repeat markers of tea. *Database*, 1-14. https://doi.org/10.1093/database/baaa013

Dutta, P. (2017). WHO encourages tea drinking for a new generation. Retrieved from https://worldteanews.com/tea-industry-news-and-features/encourages-tea-drinking-new-generation

Ellstrand, N. C., Prentice, H. C., & Hancock, J. F. (1999). Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics, 30,* 539-563. https://doi.org/10.1146/annurev.ecolsys.30.1.539

Fang, W., Meinhardt, L., Tan, H., Zhou, L., Mischke, S., & Zhang, D. (2014). Varietal identification of tea (*Camellia sinensis*) using nanofluidic array of single nucleotide polymorphism (SNP)markers. *Horticulture Research*, *1*(1403541), 1-8. https://doi.org/10.1038/hortres.2014.35

FAO. (2019). Detailed trade matrix. Retrieved from http://www.fao.org/faostat/en/#data/TM

Freeman, S., West, J., James, C., Lea, V., & Mayes, S. (2004). Isolation and characterization of highly polymorphic microsatellites in tea (*Camellia sinensis*). *Molecular Ecology Notes, 4*(3), 324–326. https://doi.org/10.1111/j.1471-8286.2004.00682.x

Furukawa, K., Sugiyama, S., Ohta, T. & Ohmido, N. (2017). Chromosome analysis of tea plant (*Camellia sinensis*) and ornamental camellia (*Camellia japonica*). *Chromosome Science, 20*(1-4), 9-15. https://doi.org/10.11352/scr.20.9

Futuyma, D. J. (1998). *Evolutionary biology* (3rd ed.). Sunderland MA: Sinauer Association. Gasura, E., Mashingaidze, B., & Mukasa, S. (2008). Genetic variability for tuber yield, quality, and virus disease complex traits in Uganda sweet potato germplasm. *African Crop Science Journal, 16*(2), 147-160. https://www.ajol.info//index.php/acsj/article/view/54355

Gerber, S., Chadoeuf, J., Gugerli, F., Lascoux, M., Buiteveld, J., Cottrell, J., Dounavi, A., Fineschi, S., Forrest, L. L., Fogelqvist, J., Goicoechea, P. G., Jensen, J. S., Salvini, D., Vendramin, G. G., Kremer, A. (2014). High rates of gene flow by pollen and seed in oak populations across Europe. *Plos One, 9*(e85130), 1-13. https://doi.org/10.1371/journal.pone.0085130

Gesimba, M., Langat, M.C., Liu, G., & Wolukau, J. N. (2005). The tea industry in Kenya: The challenges and positive developments. *Journal of Applied Sciences, 5,* 334-336. doi:10.3923/jas.2005.334.336

Graham, H. N. (1992). Green tea composition, consumption, and polyphenols chemistry. *Preventive Medicine*, 21, 334–350. https://doi.org/10.1016/0091-7435(92)90041-F

Gulati, A., Gulati, A., Ravindranath, S. D., & Chakrabarty, D. N. (1993). Economic yield losses caused by Exobasidium vexans in tea plantations. *Indian Phytopathology, 46,* 155-159. https://doi.org/10.1016/j.jksus.2020.09.008

Gupta, P., Rustgi, S., & Kulwal, P. (2005). Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology, 57*(4), 461-485. https://doi.org/10.1007/s11103-005-0257-z

Hamrick, J. L., Godt, M. J. W., & Sherman-Broyles, S. L. (1992). Factors influencing levels of genetic diversity in woody plant species. *New Forests, 6*, 95-124. https://doi.org/10.1007/BF00120641.

Harler, C. P. (1964). *The culture and marketing of tea*. New Delhi: Prentice Hall of India.

Heiss, M. L., & Heiss, J. (2007). *The story of tea: A cultural history and a drinking guide*. Berkley, CA: Ten Speed Press.

Heywood, V. (1967). Plant *Taxonomy* (2nd Ed.). London: Edwarad Arnold.

Hicks, A. (2001). Review of global tea production and the impact on industry of the Asian economic situation. *AU Journal of Technology, 5,* 1-8.

Hosseinzadeh-Colagar, A., Haghighatnia, M. J., Amiri, Z., Mohadjerani, M., & Tafrihi, M. (2016). Microsatellite (SSR) amplification by PCR usually led to polymorphic bands: Evidence which shows replication slippage occurs in extend or nascent DNA strands. *Molecular Biology Research Communication, 5*(3), 167–174.

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research, 9*(9), 868-877. https://genome.cshlp.org/content/9/9/868.full.html

Huang, H., Tong, Y., Zhang, Q. J., & Gao, L. Z. (2013). Genome size variation among and within *Camellia* species by using flow cytometric analysis. *PLoS ONE 8*(e64981), 1-14. https://doi.org/10.1371/journal.pone.0064981

Hanson, L., Mahon, K., Johnson, M. & Bennett, M. (2001). First nuclear DNA C-values for another 25 angiosperm families. *Annals of Botany, 88*(2), 851–858. https://doi.org/10.1006/anbo.2000.1325

Hazra, A., Dasgupta, N., Sengupta, C., & Das, S. (2018). Next generation crop improvement program: Progress and prospect in tea (*Camellia sinensis* (L.) O. Kuntze). *Annals of Agrarian Science, 16*(2), 128-135. https://doi.org/10.1016/j.aasci.2018.02.002

International Tea Committee. (2018). *Annual bulletin of statistics.* London: International Tea Committee.

International Tea Committee (ITC). (2019). *Annual Bulletin of Statistics*. London: ITC.

Kalinowski, S., Taper, M., & Marshall, T. (2007). Revising how the computer program

Cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology, 16*(5), 1099-1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Kamunya, S. (2010). *DNA analysis*. Kericho, Kenya: Tea Research Foundation of Kenya

Kamunya, S., Ochanda, S., Cheramgoi, E., Chalo, R., Sitienei, K., Muku, O., Kirui, W., & Bore, J. K. (2019). *Tea (Camellia sinensis (L.) O. Kuntze) production and utilization in Kenya*. Kericho, Kenya: Tea Research Institute.

Kamunya, S. M., Wachira, F. N., Pathak, R. S., Korir, R., Sharma, V., Kumar, R., Bhardwai, P., Chalo, R., Ahuja, P. S., & Sharma, R. K. (2010). Genomic mapping and testing for quantitative trait loci in tea (*Camellia sinensis* (L.) *O. Kuntze*). *Tree Genetics & Genomes, 6,* 915-929. https://doi.org/10.1007/s11295-010-0301-2

Kamunya, S. M., Wachira, F. N., Pathak, R. S., Muoki, R. C., & Sharma, R. K. (2012). Tea improvement in Kenya. In L. Chen, Z. Apostolides, & Z. Chen (Eds.), *Global tea breeding. Advanced topics in science and technology in China* (pp. 177-226). Springer.

Karak, T., & Bhagat, R. M. (2010). Trace elements in tea leaves, made tea and tea infusion: A review. *Food Research International, 43*(9), 2234-2252. https://doi.org/10.1016/j.foodres.2010.08.010

Karori, S. M., Wachira, F. N., Wanyoko, J. K., & Ngure, R. M. (2007). Antioxidant capacity of different types of tea products. *African Journal of Biotechnology, 6*(19), 2287-2296. https://doi.org/10.5897/AJB2007.000-2358

Karunarathna, K. H. T., Mewan, K. M., Weerasena, O. V., Perera, S. A., Edirisinghe, E. N., & Jayasoma, A. A. (2018). Understanding the genetic relationships and breeding patterns of Sri Lankan tea cultivars with genomic and EST-SSR markers. *Scientia Horticulturae, 240*, 72–80. https://doi.org/10.1016/j.scienta.2018.05.051

Kaundun, S., & Matsumoto, S. (2003). Development of CAPS markers based on three key genes of the phenylpropanoid pathway in Tea, *Camellia sinensis* (L.) O. Kuntze and differentiation between assamica and sinensis varieties. *Theoretical and Applied Genetics, 106*(3), 375-83. https://doi.org/10.1007/s00122-002-0999-9

Kaundun, S. S., & Park, Y. G. (2002). Genetic structure of six Korean tea populations as revealed by RAPD-PCR markers. *Crop Science, 42*(2), 594-601. https://doi.org/10.2135/cropsci2002.5940

Kaundun, S. S., Zhyvoloup, A., & Park, Y. G. (2000). Evaluation of the genetic diversity among elite tea (*Camellia sinensis* var. *sinensis*) accessions using RAPD markers. *Euphytica, 115*(1), 7-16. https://doi.org/10.1023/A:1003939120048

Kerio, L., Wachira, F. N., Kanyiri, W. J., Rotich, M. K. (2012). Characterization of anthocyanins in Kenyan teas: Extraction and identification. *Food Chemistry, 131* (1), 31-38. https://doi.org/10.1016/j.foodchem.2011.08.005

Khlestkina, E. K., Huang, X. Q., Quenum, F. J. B., Chebotar, S., Roeder, M. S., & Boerner, A. (2004). Genetic diversity in cultivated plants – loss or stability? *Theoretical and Applied Genetics, 108*(8), 1466-1472. https://doi.org/10.1007/s00122-003-1572-x

Kilel, E. C., Faraj, A. K., Wanyoko, J. K., Wachira, F. N., & Mwingirwa, V. (2013). Green tea from purple leaf coloured clones in Kenya – their quality characteristics. *Food Chemistry, 141*(2), 769-775. https://doi.org/10.1016/j.foodchem.2013.03.051

Korir, N. K., Han, J., Shangguan, L., Wang, C., Kayesh, E., Zhang, Y., & Fang, J. (2013). Plant variety and cultivar identification: Advances and prospects. *Critical Reviews in Biotechnology, 33*(2), 111-125. https://doi.org/10.3109/07388551.2012.675314

Kosman, E., & Leonard, K. J. (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular Ecology, 14*(2), 415-424. https://doi.org/10.1111/j.1365-294X.2005.02416.x

Kruglyak, S., Durrett, R. T., Schug, M. D., & Aquadro, C. F. (2000). Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. *Molecular Biology and Evolution, 17*(8), 1210-1219. https://doi.org/10.1093/oxfordjournals.molbev.a026404

Kuleung, C., Baenziger, P. S., Dweikat, I. (2004). Transferability of SSR markers among wheat, rye, and triticale. *Theoretical and Applied Genetics, 108*(6), 1147-1150. https://doi.org/10.1007/s00122-003-1532-5

Lai, Y., Tang, Q., Li, H., Chen, S., Li, P., Xu, J., Xu, Y., & Guo, X. (2016). The dark-purple tea cultivar 'Ziyan' accumulates a large amount of delphinidin-related anthocyanins. *Journal of Agriculture & Food Chemistry, 64*(13), 2719-1726. https://pubs.acs.org/doi/10.1021/acs.jafc.5b04036

Lai, J. A., Yang, W. C., & Hsiao, J. Y. (2001). An assessment of genetic relationships in cultivated tea clones and native wild tea in Taiwan using RAPD and ISSR markers. *Botanical Bulletin of Academia Sinica, 42*(2), 93-100. https://ejournal.sinica.edu.tw/bbas/content/2001/2/bot422-02.pdf

Liu, B., Sun, X., Wang, Y., Li, Y., Cheng, H., Xiong, C., & Wang, P. (2012). Genetic diversity and molecular discrimination of wild tea plants from Yunnan Province based on inter-simple sequence repeats (ISSR) markers. *African Journal of Biotechnology, 11*(53), 11566-11574.

Lu, Y. (1974). *The classic of tea*. Introduced and Transl. by F. Ross Carpenter and H. Demi (Boston, MA: Little Brown & Co.).

Lu, H. Y., Zhang J. P., Yang Y. M., Yang X. Y., Xu B. Q., Yang W. Z., Tong, T., Jin, S., Shen, C., Rao, H., Li, X., Lu, H., Fuller, D. Q., Wang, L., Wang, C., Xu, D., & Wu, N. (2016). Earliest tea as evidence for one branch of the silk road across the Tibetan Plateau. *Scientific Reports, 6*(18955), 1-16. https://doi.org/10.1038/srep18955

Ma, J. Q., Zhou, Y. H., Ma, C. L., Yao, M. Z., Jin, J. Q., Wang, X. C., & Chen, L. (2010). Identification and characterization of 74 novel polymorphic EST-SSR markers in the tea plant, *Camellia sinensis* (Theaceae). *American Journal of Botany, 9*(12), e153-e156. https://doi.org/10.3732/ajb.1000376

Magoma, G. N., Wachira, F. N., Obanda, M., Imbuga M., & Agong, S. G. (2000). The use of catechins as biochemical markers in diversity studies of tea (*Camellia* sinensis). *Genetic Resources and Crops Evolution, 47*(2), 107-114. https://doi.org/10.1023/A:1008772902917

Mahmood, T., Akhtar, N., & Khan, B. A. (2010). The morphology, characteristics, and medicinal properties of *Camellia* sinensis' tea. *Journal of Medicinal Plants Research, 4*(19), 2028-2033. https://doi.org/10.5897/JMPR10.010

Matsumoto, S., Kiriiwa, Y., & Takeda, Y. (2002). Differentiation of Japanese green tea cultivars as revealed by RFLP analysis of phenylalanine ammonia-lyase DNA. *Theoretical and Applied Genetics, 104*(6-7), 998-1002. https://doi.org/10.1007/s00122-001-0806-z

Matsumoto, S., Takeuchi, A., Hayatsu, M., & Kondo, S. (1994). Molecular cloning of phenyllanine ammonia-lyase cDNA and classification of varieties and cultivars of tea plants (*Camellia sinensis*) using the tea PAL cDNA probe. *Theoretical and Applied Genetics, 89*(6), 671-675. https://doi.org/10.1007/BF00223703

Meegahakumbura M. K., Wambulwa M. C., Thapa K. K., Li M. M., Möller M., Xu J. C., Yang, J. B., Liu, B. Y., Ranjitkar, S., Liu, J., Li, D. Z., & Gao, L. M. (2016). Indications for three independent domestication events for tea plant (*Camellia sinensis* (L.) O. Kuntze) and new insights into the origin of tea germplasm in China and India revealed by nuclear microsatellites. *PLoS ONE 11*(5), 1-11. https://doi.org/10.1371/journal.pone.0155369

Ming, T. L., & Bartholomew, B. (2007). Theaceae. In Z. Y., Wu, P. H., Raven, & Hong, D. Y. (Eds.), *Flora of China (Hippocastanaceae through Theaceae)* (pp. 366-478). St. Louis: Science Press, Beijing and Missouri Botanical Garden Press.

Mingsheng, W., Jia, X., Tian, L., & Lv, B. (2006). Rapid and reliable purity identification of F1 hybrids of maize (*Zea mays* L.) using SSR markers. *Molecular Plant Breeding, 4*(3), 381-384.

Mondal, T. K. (2002). Assessment of genetic diversity of tea (*Camellia sinensis* (L.) O. Kuntze) by inter-simple sequence repeat polymerase chain reaction. *Euphytica, 128*(3), 307-315. https://doi.org/10.1023/A:1021212419811

Mondal, T. K. (2011) Camellia. In: Kole C (ed) *Wild crop relatives: Genomic and breeding resources* (pp. 15-40). Springer-Verlag, Heidelberg,

Mondal, T. K., Bhattacharya, A., Laxmikumaran, M., & Ahuja, P. S. (2004). Recent advances of tea (*Camellia sinensis*) biotechnology. *Plant Cell, Tissue and Organ Culture, 76*(3), 195-254. https://doi.org/10.1023/B:TICU.0000009254.87882.71

Mukhopadhyay, M., Mondal, T. K., & Chand, P. K. (2016). Biotechnological advances in tea (*Camellia sinensis* [L.] O. Kuntze): A review. *Plant Cell Reports, 35*, 255-287. https://doi.org/10.1007/s00299-015-1884-8

Muoki, C. R., Maritim, T. K., Oluoch, W. A., Kamunya, S. M., & Bore, J. K. (2020). Combating climate change in the tea industry. *Frontiers in Plant Science, 11*(339), 1-10. https://doi.org/10.3389/fpls.2020.00339

Murr, L. A., Hauck, B., Winters, A., Heald, J., Lloyd, A. J., Chakraborty, U., & Chakraborty, B. N. (2015). The development of tea blister caused by *Exobasidium vexans* in tea (*Camellia sinensis*) correlates with the reduced accumulation of some antimicrobial metabolites and the defence signals salicylic and jasmonic acids. *Plant Pathology, 64*(6), 1471-1483. https://doi.org/10.1111/ppa.12364

Mwangi, M. C. (2016). The causes of high cost of tea production and sustainability of the tea subsector in Kenya. *International Journal of Science and Research, 5*(9), 1186-1189. doi:10.21275/ART20161721

Nagy, S., Poczai, P., Cernak, I., Gorji, A. M., Hegedus, G., & Taller, J. (2012). PICcalc: An online program to calculate polymorphic information content for molecular genetic studies. *Biochemical Genetics, 50*(9-10), 570-672. https://doi.org/10.1007/s10528-012-9509-1

Nandakumar, N., Singh, A. K. Sharma, R., Mohapatra, K., & Prabhu, V. (2004). Molecular fingerprinting of hybrids and assessment of genetic purity of hybrid seeds in rice using microsatellite markers. *Euphytica, 136*(3), 257-264. https://doi.org/10.1023/B:EUPH.0000032706.92360.c6

Navajas, M., & Fenton B. (2000). The application of molecular markers in the study of diversity in acarology: A review. *Experimental and Applied Acarology, 24*(10-11), 751-774. https://doi.org/10.1023/A:1006497906793

Ndege, P. O. (2021). *"All time is tea time": The prospects and challenges of the global tea industry*. Retrieved from https://www.africamultiple.uni-bayreuth.de/en/news/2021/2021-05-17_tea/index.html

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of National Academy of Sciences U.S.A., 70,* 3321–3323.

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America, 76*(10), 5269-5273. https://doi.org/10.1073/pnas.76.10.5269

National Institute of Standards and Technology [NIST]. (2016). *Shannon diversity index*. Retrieved from https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/shannon.htm

Niu, S., Song, Q., Koiwa, H., Qiao, D., Zhao, D., Chen, Z., Liu, X., & Wen, X. (2019). Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. *BMC Plant Biology, 19*(328), 1-12. https://doi.org/10.1186/s12870-019-1917-5

Oliveira, E., Padua, J., Zucchi, M., Vencovsky, R. and Vieira, M. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology, 29*(2), 294-307. http://dx.doi.org/10.1590/S1415-47572006000200018

Olson, R., Francis, C., & Kaffka, S. (1995). *Exploring the role of diversity in sustainable agriculture*. New York: American Society of Agronomy, Inc.

Onduru, D. D., Jager, A. D., Hiller, S. R/. & Bosch, R. V. (2012). Sustainability of smallholder tea production in developing countries: Learning experiences from farmer field schools in Kenya. *International Journal of Development and Sustainability, 1*(3), 714-742.

Onsando, J. M., Wargo, P., & Wando, S. W. (1997). Distribution, severity, and spread of Armillaria root disease in Kenya tea plantations. *Plant Disease, 81*, 133-137. https://doi.org/10.1094/PDIS.1997.81.2.133

Pandey, A. K., Sinniah, G. D., Babu, A., & Tanti, A. (2021). How the global tea industry copes with fungal diseases – Challenges and opportunities. *Plant Disease, 105*(7), 1868-1879. https://doi.org/10.1094/PDIS-09-20-1945-FE

Parthiban, S., Govindaraj, P., & Senthilkumar, S. (2018). Comparison of relative efficiency of genomic SSR and EST-SSR markers in estimating genetic diversity in sugarcane. *3 Biotech, 8*(3), 144-149. https://doi.org/10.1007/s13205-018-1172-8

Paul, S., Wachira, F. N., Powell, W., & Waugh, R. (1997). Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. *Theoretical and Applied Genetics, 94*(2), 255-263. https://doi.org/10.1007/s001220050408

Pavel, A. B., & Vasile, C. I. (2012). PyElph - A software tool for gel images analysis and phylogenetics. *BMC Bioinformatics, 13***(**9), 1-6. https://doi.org/10.1186/1471-2105-13-9

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics, 28*(19), 2537-2539. https://doi.org/10.1093/bioinformatics/bts460

Pereira, A. G., Garcia-Perez, P., Cassani, L., Chamorro, F., Cao, H., Barba, F. J., Simal-Gandara, J., & Prieto, M. A. (2022). *Camellia japonica:* A phytochemical perspective and current applications facing its industrial exploitation. *Food Chemistry: X, 13,* 1-12. https://doi.org/10.1016/j.fochx.2022.100258

Preedy, V. R. (2012). *Tea in health and disease prevention*. New York: Academic Press.

Prince, L. M., & Parks, C. R. (2001). Phylogenetic relationships of Theaceae inferred from chloroplast DNA sequence data. *American Journal of Botany, 88*(12), 2309-2320. https://www.jstor.org/stable/3558391

Pritchard, J. K., Stephens, M., Rosenberg, N. A., & Donnelly P. (2000). Association mapping in structured populations. *American Journal of Human Genetics, 67,* 170-181. https://doi.org/10.1086/302959

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution, 4,* 4399-4428. https://doi.org/10.1002/ece3.1305

Radhakrishnan, B., & Baby, U. I. (2004). Economic threshold level for blister blight of tea. *Indian Phytopathology, 57,* 195-196.

RCMRD Geoportal. (2015). *Kenya county boundaries*. Retrieved from http://geoportal.rcmrd.org/layers/servir%3Akenya_county_boundary

Reyes-Valdés, M. H. (2013). Informativeness of microsatellite markers. *Methods in Molecular Biology, 1006,* 259-270. https://doi.org/10.1007/978-1-62703-389-3_18

Rhouma, H. B., Taski-Ajdukovic, K., Zitouna, N., Sdouga, D., Milic, D., & Trifi-Farah, N. (2017). Assessment of the genetic variation in alfalfa genotypes using SRAP markers for breeding purposes. *Chilean Journal of Agricultural Research, 77*(4), 332-340. http://dx.doi.org/10.4067/S0718-58392017000400332

Rider, M. (2022). Leading tea exporters worldwide 2020. Retrieved from https://www.statista.com/statistics/264189/main-export-countries-for-tea-worldwide/

Rocha, E. A., Paiva, L. V., de Carvalho, H. H., & Guimarães, C. T. (2010). Molecular characterization and genetic diversity of potato cultivars using SSR and RAPD markers. *Crop Breeding and Applied Biotechnology, 10*(3), 1-8. https://doi.org/10.1590/S1984-70332010000300004

Rozen, S., & Skaletsky, H. J. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology, 132*, 365–386. https://doi.org/10.1385/1-59259-192-2:365

SantaLucia, J. 2007. Physical principles and visual-OMP software for optimal PCR design. *Methods Molecular Biology, 402,* 3-34. https://doi.org/10.1007/978-1-59745-528-2_1

Sharma, R. K., Bhardwaj, P., Negi, R., Mohapatra, T., & Ahuja, P. S. (2009). Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). *BMC Plant Biology, 9*(53), 1-12.

https://doi.org/10.1186/1471-2229-9-53

Sheidai, M., Jahanbakht, H. & Sofi-Siyavash, P. (2004). Cytogenetic study of various types of tea (*Camellia sinensis*) cultivars in Iran. *Iranian Journal of Science & Technology Transaction A-Science, 28*(A1), 33-42 (2004).

Stoeckle, M., Gamble, C., Kirpekar, R., Young, G., Ahmed, S., & Little, P. (2011). Commercial Teas highlight plant DNA barcode identification successes and obstacles. *Scientific Reports, 1*(42), 1-7. https://doi.org/10.1038/srep00042

Su, T., Wang, C., Lin, K., Chang, M., Kang, R., Nguyen, K. M., & Nguyen, H. (2017). Genetic diversity of a novel oil crop, *Camellia brevistyla*, revealed by ISSR DNA markers. *Horticultural Science and Technology,* 588-598. https://doi.org/10.12972/kjhst.20170063

Syahbudin, A., Widyastuti, A., Masruri, N. W., & Meinata, A. (2019). Morphological classification of tea clones (Camellia sinensis, Theaceae) at the Mount Lawu Forest, East Java, Indonesia. *Earth and Environmental Science, 394,* 1-12. doi:10.1088/1755-1315/394/1/012014

Taheri, S., Abdullah, T. L., Yusop, M. R., Hanafi, M. M., Sahebi, M., Azizi, P., & Shamshiri, R. R. (2018). Mining and development of novel SSR markers using next generation (NGS) data in plants. *Molecules, 23*(2), 399-406. https://doi.org/10.3390/molecules23020399

Tan, L. Q., Wang, L. Y., Wei, K., Zhang, C. C., Wu, L. Y., Qi, G. N., Cheng, H., Zhang, Q., Cui, Q. M., & Liang, J. B. (2013). Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). *PLoS ONE, 8* (e81611), 1-13. https://doi.org/10.1371/journal.pone.0081611

Tanaka, J., Taniguchi, F., Hirai, N. & Yamaguchi, S. (2006). Estimation of the genome size of tea (*Camellia sinensis*), Camellia (*C. japonica*), and their interspecific hybrids by flow cytometry. *Tea Research Journal, 1,* 1–7. https://doi.org/10.5979/cha.2006.1

Taski-Ajdukovic, K., Nagl, N., Milic, D., Katic, S., & Zoric, M. (2014). Genetic variation and relationship of alfalfa populations and their progenies based on RAPD markers. *Central European Journal of Biology, 9*(8), 768-776. https://doi.org/10.2478/s11535-014-0307-0

Tea Board of Kenya. (2010). *Annual report and accounts 2009/2010*. Nairobi: Tea Board of Kenya.

Tea Research Foundation of Kenya [TRFK]. (2012). *Annual technical report: Molecular investigations*. Kericho: TRFK.

Tea Research Institute [TRI]. (2019). *Annual technical report for 2019*. Kericho: TRI.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., & McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research, 11*(8), 1441-1452. http://www.genome.org/cgi/doi/10.1101/gr.184001

Tessier, C., David, J., This, P., Boursiquot, J. M., & Charrier, A. (1999). Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theoretical and Applied Genetics, 98*(1), 171-177. https://doi.org/10.1007/s001220051054

Thuravaki, B., Ranatunga, M. A. B., Kottawa-Arachchi, J. D., & Sumanasinghe, V. A. (2017). Characterization of new tea (Camellia sinensis L.) hybrid progenies based on morphological traits. *International Journal of Tea Science, 13*(1&2), 01-09. https://doi.org/10.20425/ijts.v13i01-02.9980

Tsykun, T., Rellstab, C., Dutech, C., Sipos, G., & Prospero, S. (2017). Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*. *Heredity, 119*(5), 371-380. https://doi.org/10.1038/hdy.2017.48

Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genic microsatellite markers in plants: Features and applications. *Trends in Biotechnology, 23*(1), 48-55. https://doi.org/10.1016/j.tibtech.2004.11.005

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Freijters, A., Pot, J., Peleman, J., Kuiper, M., & Zabeau, M. (1995). AFLP: A new techniques for DNA fingerprinting. *Nucleic Acids Research, 23*(21), 4407-4414. https://doi.org/10.1093/nar/23.21.4407

Wachira, F. N. (2002). Genetic mapping of tea: A review of achievements and opportunities. *Tea, 23*(2), 91-102.

Wachira, F., & Kamunya, S. (2005). Pseudo-self-incompatibility in some tea clones (*Camellia sinensis* (L.) O. Kuntze). *Journal of Horticultural Science and Biotechnology, 80,* 716-720. https://doi.org/10.1080/14620316.2005.11512004

Wachira, F. N., & Kamunya, S. K. (2017). *Evidence of late acting self incompatibility in tea.* Retrieved from http://www.ocha.net/english/conference2/pdf/2004/files/PROC/Pr-O-12.pdf

Wachira, F. N., Kamunya, S., Karori, S., Chalo, R., & Maritim, T. (2013). The tea plants: Botanical aspects. In V. R. Preedy (Ed.), *Tea in health and disease prevention* (pp. 7-24). Academic Press.

Wachira, F. N., Powell, W., Waugh, R. (1997). An assessment of genetic diversity among *Camellia* sinensis L. (cultivated tea) and its wild relatives based on randomly amplified polymorphic DNA and organelle-specific STS. *Heredity, 78*(6), 603-611. https://doi.org/10.1038/hdy.1997.99

Wachira, F., Tanake, J., & Takeda, Y. (2001). Genetic Variation and differentiation of tea (*Camellia sinensis*) germplasm revealed by RAPF and AFLP variation. *Journal of Horticultural Science & Biotechnology, 76*(5), 557-563. https://doi.org/10.1080/14620316.2001.11511410

Wachira, F. N., Waugh, R., Hackett, C.A., & Powell, W. (1995). Detection of genetic diversity in tea (*Camellia sinensis*) using RAPD markers. *Genome, 38*(2), 201-210. https://doi.org/10.1139/g95-025

Wambulwa, M., Meegahakumbura, M., Chalo, R., Kamunya, S., Muchugu, A., Xu, J., Liu, J., Li, D., & Gao, L. (2016). Nuclear microsatellites reveal the genetic architecture and breeding history of tea germplasm of East Africa. *Tree Genetics & Genomes, 12*(1), 11-21. https://doi.org/10.1007/s11295-015-0963-x

Wambulwa, M. C., Meegahakumbura, M. K., Kamunya, S., Muchugi, A., Möller, M., Liu, J., Xu, J. C., Li, D. Z., & Gao, L. M. (2017). Multiple origins and a narrow gene pool characterise the African tea germplasm: Concordant patterns revealed by nuclear and plastid DNA markers. *Scientific Reports, 7*(4053), 4053-4061. https://doi.org/10.1038/s41598-017-04228-0

Wei, K., Wang, L., Zhou, J., He, W., Zong, J., Jiang, Y., & Cheng, H. (2012). Comparison of catechins and purine alkaloids in albino and normal green tea cultivars (*Camellia sinensis* L.) by HPLC. *Food Chemistry,* 130(1), 720–724. https://doi.org/10.1016/j.foodchem.2011.07.092

Weising, K., Nybom, H, Wolff, K., & Kahl, G. (2005). Detecting DNA variation by molecular markers. In H. Nybom, K. Weising & M. Pfenninger (Eds.), *DNA fingerprinting in Plants: Principles, methods, and applications* (pp. 21-73). New York, NY: Taylor & Francis Group.

Wight, W. (1959). Nomenclature and classification of the tea plant. *Nature, 183,* 1726-1728. https://doi.org/10.1038/1831726a0

Williams, J. G. K., Kubelik, A. R. K., Livak, J. L., Rafalski, J. A., & Tingey, S. V. (1990). DNA polymorphisms amplified by random primers are useful as genetic markers. *Nucleic Acids Research, 18*(22), 6531-6535. https://doi.org/10.1093/nar/18.22.6531

Willson, K. C. & Clifford, M. N. (1992). *Tea cultivation to consumption*. London: Chapman and Hall.

Wu, H., Chen, D., Li, J., Yu, B., Qiao, X., Huang, H., & He, Y. (2013). De novo characterization of leaf transcriptome using 454 sequencing and development of EST-SSR markers in tea (*Camellia sinensis*). *Plant Molecular Biology Reporter, 31*(3), 524-538. https://doi.org/10.1007/s11105-012-0519-2

Xia, E., Tong, W., Wu, Q., Wei, S., Zhao, J., Zhang, Z., Wei, L., & Wan, X. (2020). Tea plant genomics: Achievements, challenges and perspectives. *Horticulture Research, 7*(7), 1-19. https://doi.org/10.1038/s41438-019-0225-4

Xia, J., Liu, Y., Yao, S., Li, M., Zhu, M., Huang, K., Gao, L., & Xia, T. (2017).

Characterization and expression profiling of *Camellia sinensis* cinnamate 4-hydroxylase genes in phenylpropanoid pathways. *Genes, 8*(8), 193-198. https://doi.org/10.3390/genes8080193

Xu, Y., Qiao, F., & Huang, J. (2022). Black tea markets worldwide: Are they integrated? *Journal of Integrative Agriculture, 21*(2), 552-565. https://doi.org/10.1016/S2095-3119(21)63850-9

Yamamoto, T., Juneja, L. R, Chu, D. C., & Kim, M. (1997). *Chemistry and application of green tea*. Boca Raton, FL: CRC Press.

Yamanishi T. (1995). Special issue on tea. *Food Reviews International, 11*, 371–546.

Yamashita, H., Katai, H., Kawaguchi, L., Nagano, A. I., Nakamura, Y., Morita, A., Ikka, T. (2019). Analyses of single nucleotide polymorphisms identified by ddRAD-seq reveal genetic structure of tea germplasm and Japanese landraces for tea breeding. *PLoSONE 14*(8):e0220981, 1-15. .https://doi.org/10.1371/journal.pone.0220981

Yao, M., Huang, H., Yu, J., & Chen, L. (2005). Analysis on applicability of ISSR in molecular identification and relationship investigation of tea cultivars. *Journal of Tea Science, 25*(2), 153-157. Doi: 10.13305/j.cnki.jts.2005.02.013

Yao, M. Z., Ma, C. L., Qiao, T. T., Jin, J., & Chen, L. (2012). Diversity distribution and population structure of tea germplasms in China revealed by EST-SSR markers. *Tree Genetics & Genomes, 8*(1), 205-220. https://doi.org/10.1007/s11295-011-0433-z

Yashitola, J., Thirumurugan, T., Sundaram, R., Naseerullah, M., & Ramesha, M., Sarma, N. P., & Sonti, P. R. (2002). Assessment of purity of rice hybrids using microsatellite and STS markers. *Crop Science, 42*(4), 1369-1373. https://doi.org/10.2135/cropsci2002.1369

Yeh, F. C., Boyle, R., Yang, R. C., Ye, Z., Mao, J. X., & Yeh, D. (1999). *POPGENE version 1.32. Computer program and documentation*. Retrieved from http://www.ualberta.ca/~fyeh/popgene.html

Yeeh, Y., Soon, S. K., & Myong, G. C. (1996). Evaluations of the natural monument populations of *Camellia* japonica (Theaceae) in Korea based on allozyme studies. *Botanical Bulletin of Academia Sinica, 37,* 141 - 146. Retrieved from https://ejournal.sinica.edu.tw/bbas/content/1996/2/bot372-08.html

Zhang, K., Wu, Z., Tang, D., Lv, C., Luo, K., Zhao, Y., Liu, X., Huang, Y., & Wang, J. (2016). Development and identification of SSR markers associated with starch properties and β-carotene content in the storage root of sweet potato (*Ipomoea batatas* L.). *Frontiers in Plant Science, 7*(223), 1-21. https://doi.org/10.3389/fpls.2016.00223

Zhang, C.C., Wang, L., Wei, L., Wu, H., Li, F., Zhang, F., & Cheng, D. (2016). Transcriptome analysis reveals self-incompatibility in the tea plant (Camellia sinensis) might be under gametophytic control. *BMC Genomics, 17*(1), 359-367. https://doi.org/10.1186/s12864-016-2703-5

Zhang, Y., Zhang, X., Chen, X., Sun, W., & Li, J. (2018). Genetic diversity and structure of tea plant in Qinba area in China by three types of molecular markers. *Hereditas, 155*(22), 1-12. https://doi.org/10.1186/s41065-018-0058-4

Zhao, D., Yang, J., Yang, S., Keto, K., & Luo, J. (2014). Genetic diversity and domestication origin of tea plant *Camellia taliensis* (Theaceae) as revealed by microsatellite markers. *BMC Plant Biology 14*(14), 1-12. https://doi.org/10.1186/1471-2229-14-14

Zhen, Y., Chen, Z., Cheng, S., & Chen, M. (2005). *Tea: bioactivity and therapeutic potential*. New York: Taylor & Francis.

Zhou, Q., Li, H., Hoang, T. X., & Ruan, X. (2019). Genetic diversity and relationship of dongting Bilouchun tea germplasm in Suzhou revealed by SSR markers. *Pakistan Journal of Botany, 51*(3), 1-8.Retrieved from http://pakbs.org/pjbot/papers/1548240029.pdf

Zong, J., Zhaol, T., Ma, Q., Liang, L., Wang, G. (2015). Assessment of genetic diversity and population genetic structure of *Corylus mandshurica* in China using SSR markers. *PLoS ONE, 10*(9), 1-12. https://doi.org/10.1371/journal.pone.0137528

APPENDICES

Appendix I: Characteristics of the EST-SSR and adapted microsatellite markers used to study genetic diversity among interspecific hybrids of tea

Code	Sequences (5' to 3')	Length (bp)	GC%	Target motif	Melting temperature Tm (°C)	Mol. Weight (g/mol)	Product size range (bp)	Reference
CamJap 1	F_ AACAGCAGCAACAGCAACAA	20	45	(CCG)7	64.7	6186.73	280	Novel EST
	R_TCCATCCAATACTGCAAGTCC	21	47.6		63.8	6390.82		
CamTal	F_CCTTCGCTCACCATTCTTTC	20	50.0	(TTC)6	59.8	6010.49	168	Novel EST
	R_TGTAGCCCATTCCCTTTGTC	20	50.0		59.9	6090.55		
CamJap 2	F_CCTTGTCTGTAATGCCTCTCAA	22	45.5	(CAG)4	59.4	6717.04	259	Novel EST
	R_TGTTGTTGTTGCCTGTTGGT	20	45.0		60.0	6247.64		
CamJap 3	F_AGCCAAGAAGATGTCCTCCA	20	50.0	(AAC)4	59.8	6175.68	176	Novel EST
	R_CATCACCACCAACTCCATCA	20	50.0		60.4	6015.56		
CamJap 4	F_CACGATTCCTCTCAGCAACA	20	50.0	(AAC)5	60.0	6086.6	184	Novel EST
	R_GACTTCCATCGGAATCCTCA	20	50.0		60.0	6117.61		
TM 134	F-TTCCGTGACTGATTTATGTG	20		(CAT)8	56	6128.9	221-251	Wambulwa *et al*. (2016)
	R-TTGAGACTCGGGGTTTT	17	47.1			5247.4		
TM 179	F-GTCCCAGAAATCATAACG	18	44.4	(TGA)8	58	5476.1	135-162	Wambulwa *et al*. (2016)
	R-CGACAAGGGATTAGCAG	18	44.4			5476.1		
TM 197	F-GAGGAGCATTAGCATCTT	18	44.4	(AGG)7	59	5538.3	118-142	

	R-GGACCAGTACGAGTAGC	17	58.8			5243.9		Wambulwa *et al*. (2016)
TM 203	F-AGAGCTTCTCAACAACCC	18	50.0	(GAT)9	57	5412.1	165-200	Wambulwa *et al*. (2016)
	R-ATGGAGCATACTACTCACTT	20	40.0			6075.6		
TM 51	F-AATCATGCCCAAGGACATTC	20	45.0	(GGT)6	60	6069.4	168-189	Wambulwa *et al*. (2016)
	R-CAACCACTACCCATTTCACT	20	45.0			5940.5		
TM 58	F-CATTATCCCTTTCCTTGTCCA	21	42.9	(TCA)6	61	6258.0	225-252	Wambulwa *et al*. (2016)
	R-GGAGGGAGTAGGAGGTCT	18	61.1			5684.1		
TUGMS 2-135	F-ATGCTAGCCATGGCAATACC	20	50.0	(GAA)8	56	6085.4	238-289	Wambulwa *et al*. (2016)
	R-CACACTGCACATGATGGTGA	20	50.0			6125.4		
TUGMS 2-157	F-CCCATGGTCTATTTCGCTGT	20	50.0	(CCA)16	53.5	6049.8	165-186	Wambulwa *et al*. (2016)
	R-CCAGAGATGGACCTGACACA	20	55.0			6119.2		
A37	F-TCTGCCCTTCCCTAAATC	18	50.0	(AAG)9	54	5345.4	170-182	Wambulwa *et al*. (2016)
	R-ATGTTTGGTCTCGGTTGTT	19	42.1			5846.9		
A 47	F-TCCCTACAAACCCTAACCG	19	52.6	(GCC) 5	61	5661.2	171-201	Wambulwa *et al*. (2016)
	R-GAGCAGCATCAGAGTCACGT	20	55.0			6150.3		
Camsin M1	F_GAATCAGGACATTATAGGAATTAA	24	29.2	(GT)16	48.4	7432.7	280–300	Freeman *et al*. (2004)
	R_GGC CGA ATG TTG TCT TTT GT	20	45.0		53.8	6144.9		
	F_CCT CTG GGT GTC CTA CAC CT	20	60.0	(GT)17	52.5	6019.8	240–260	

Camsin M2	R_AAA GCC TTG ATG CCT TTC G	19	47.4		54.2	5778.7		Freeman *et al*. (2004)
Camsin M3	F_GGT GTG GTG TTT TGA AGA AA	20	40.0	(CA)18	49.6	6267.0	190–210	Freeman *et al*. (2004)
	R_TGT TAA GCC GCT TCA ATG C	19	47.4		53.7	5778.7		
Camsin M4	F_ACATTCAAGCANTCCACATATGTGAAA	27	35.2	(GA)19	59.5	8240.0	358–370	Freeman *et al*. (2004)
	R_CCTGNTGCAGGACTGTCTATAGATGA	26	48.1		58.5	8005.8		
Camsin M5	F_AAACTTCAACAACCAGCTCTGGTA	24	41.7	(GT)15(GA)8	55.9	7289.6	170–205	Freeman *et al*. (2004)
	R_ATTATAGGATGCAAACAGGCATGA	24	37.5		56.4	7433.7		

Appendix II: Composition of the PCR master mix

Component	**Final volume**
Reaction buffer (1x)	1µl
MgCl2 (2mM)	1µl
Forward primer (0.5 µM)	0.5µl
Reverse primer (0.5 µM)	0.5µl
dNTP mix (0.2mM)	0.5µl
Taq DNA polymerase (5U/ µl)	0.2µl
DNA template (20ng/µl)	2µl
DH$_2$0	4.3µl
Total volume	**10µl**

Appendix III: PCR conditions used

Step	No. of cycles	Temperature (°C)	Time
Initial denaturation	1	94	4 min.
Denaturation	35	94	30 sec.
Annealing		55	1 min.
Extension		72	30 sec.
Final extension	1	72	7 min.

Appendix IV: The presence and absence of bands generated from 88 genotypes with 8 SSR primer pairs (1: presence; 0: absence; ?: missing data)

		Camsin				TM		
Genotype	**Camsin M5**	**M2**	**Camjap A1**	**Camjap A4**	**TM 51**	**134**	**A37**	**A47**
C.	? ?	0 0 1 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 1 0	0 0 0 0	0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
japonica	? ? ? ? ?	1 1	1 0 0 0 1 1	0 0 0 1 1	0 0 0	0 1	0 0	1 0 0 0 0
C.	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 1 0 1	0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
sasanqua	0 1 0 0 0 0 1	1 1	0 0 0 0 1 1	0 0 0 1 1	0 1 0	0 1	0 0	1 0 1 0 0
C.	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1	0 0 1 0 1	0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 1 1 0	0 0 0 1	1 0 1 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
brevistyla	0 1 0 0 0 0 1	0 0	1 0 0 0 0 1	0 0 0 0 1	0 1 0	0 1	0 0	1 1 0 0 0
	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1	0 0 0 1 0	0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1	0 0 0 1	0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
C. oleifera	0 1 0 0 0 0 1	1 0	0 0 0 0 1 1	0 0 0 1 1	0 1 0	0 1	0 0	1 1 0 0 0
	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1	0 0 0 1 0	0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1	0 0 0 0	? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
C. kissi	0 1 0 0 0 0 1	1 0	0 0 0 0 1 1	0 0 0 1 1	0 1 0	0 1	? ?	1 1 1 1 0
C.								
irrawandie	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0	0 0 0 1 0	0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 1 0	0 0 0 0 0 0 0 0 0 1 0 0 1 0 0	0 0 0 0 0 0 0 1	0 0 0 1	0 1 0 0 1	0 0 0 0 0 1 0 0 0 0 0 0 0 0 1
nsis	1 1 0 0 0 0 1	1 0	0 1 0 1 0 1	0 1 1 0 1	0 1 0	0 1	0 0	1 1 0 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 1 0	0 0 0 0 0 0 0 0 1 0 1 0 1 0 0 0 1 0	0 0 0 0 1 1 1 1 1 1 0 0 1 0 0	0 1 1 0 0 0 0 1	0 0 0 1	0 1 0 0 1	0 1 1 1 0 1 0 1 0 1 0 0 1 0 1
306/1	1 1 0 0 0 0 1	0 0	0 1 0 1 0 1	0 1 1 0 1	0 1 0	0 1	0 0	1 1 0 0 0

TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1	0 0 1 1	0 1 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
306/2	1 1 0 0 0 0 1	0 0	1 1 0 0 0 1	0 0 0 0 1	0 0 0	0 1	0 0	0 0 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 1	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0	0 1 0 0 0 0 1 1	0 0 1 1	0 1 0 0 1	0 0 0 0 0 1 0 1 0 1 0 0 0 0 1
306/3	1 1 0 0 0 0 1	0 0	0 1 0 1 0 1	0 0 1 0 1	0 1 0	0 1	0 0	0 1 0 0 0
TRFK	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0 1 0 0 1	0 0 1 0 0	0 0 0 0 0 0 0 0 1 0 0 0 1 1 0 0 1 1	0 0 0 0 0 1 0 1 0 1 0 0 1 0 1	0 1 0 1 0 0 1 1	0 0 0 0	0 1 0 0 1	0 1 0 1 0 1 1 1 0 1 0 0 1 0 1
306/4	0 1 0 0 0 0 1	0 0	0 1 0 1 0 1	0 1 0 1 1	0 1 0	0 1	0 0	0 1 0 0 0
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 1 1	0 0 0 1	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRFK 73/1	0 1 0 0 0 0 1	0 0	1 1 0 0 1 1	0 1 0 1 1	0 1 0	0 1	1 0	1 1 0 0 0
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 1 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 1	0 0 1 1	0 1 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
TRFK 73/2	0 1 0 0 0 0 1	0 0	0 1 0 0 1 1	0 1 0 1 1	1 0 1	0 1	0 0	1 1 0 0 0
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0	1 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 1 0 1	0 0 0 0 0 1 0 0	0 0 0 1	0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
TRFK 73/3	0 1 1 0 0 0 1	0 0	0 1 0 1 1 1	1 0 0 1 1	1 0 1	0 1	0 0	1 1 0 0 0
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 0 1 0 0	0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 1 0 1	0 0 0 1 0 0 0 1	0 0 1 1	0 0 0 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
TRFK 73/4	0 1 0 0 0 0 1	0 0	0 1 1 0 1 1	1 0 0 1 1	1 0 1	0 1	0 0	1 1 0 0 0
	0 0 0 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0	0 0 0 1 0 1 1 0 1 0 1 1 1 0 0 1 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 0	0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
TRFK 73/5	1 1 0 0 0 0 1	1 0	0 0 0 0 0 0	0 0 0 0 1	0 0 1	0 1	0 0	0 0 0 0 0
	0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0	0 1 0 1 0 1 1 0 1 0 1 1 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1	0 0 0 1	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRFK 91/2	0 1 1 0 0 0 1	1 0	0 0 0 0 0 0	0 0 1 0 1	0 1 1	0 1	1 0	1 1 0 0 0
	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0	0 0 1 0 0 1 1 0 0 1 1 0 0 0 0	0 0 0 0 0 0 0 0	0 1 0 0	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
TRFK 83/1	1 0 1 0 0 0 1	0 0	0 0 0 0 0 0	0 0 0 0 1	0 0 1	0 0	1 0	0 1 0 1 0
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRFK 14/1	0 1 1 0 0 0 1	1 0	0 1 0 1 0 1	0 0 0 0 1	0 0 1	0 1	1 0	0 0 1 0 0
TRFK	0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 1	0 0 0 0 0	0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0	0 0 0 0 0 0 0 0 0 1 0 0 1 0 0	0 0 1 0 0 0 0 1	0 0 1 1	1 0 1 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
301/1	0 0 1 0 0 0 1	1 0	0 1 0 1 1 1	0 0 0 1 1	0 1 1	0 1	0 0	1 1 0 0 0
TRFK	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 1	? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
31/11	0 0 1 0 0 0 1	1 0	0 0 1 0 1 1	0 0 0 1 1	0 0 1	0 1	? ?	1 0 0 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0	0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
31/32	0 1 0 0 1 0 1	1 0	0 0 0 0 0 0	0 0 1 0 0	0 1 1	0 1	1 0	0 0 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 1	0 1 1 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
31/33	0 0 1 1 0 0 1	1 0	0 1 0 0 1 1	0 0 0 1 1	0 1 1	0 1	0 0	1 1 0 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 1	0 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
31/34	0 0 0 1 1 0 1	1 0	0 0 0 0 1 1	0 0 0 1 1	0 1 1	0 1	1 0	0 1 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 1 0	0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
31/35	0 1 0 0 1 0 1	1 0	0 0 0 1 0 1	0 0 0 0 1	0 0 1	0 1	1 0	0 1 0 1 0

TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 1 1	0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 1	0 0 1 0 1	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
31/36	0 1 0 0 1 0 1	0 0	0 0 0 1 0 1	0 0 0 0 1	0 0 1	0 1	0 0	? ? ? ?
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0	0 0 0 1 1	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 1 0 0	0 0 0 0 0 0 0 1	0 0 0 1	0 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
31/38	1 0 0 1 0 0 1	0 0	0 0 0 1 0 1	0 1 0 1 0	1 0 1	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 1 0	0 1 1 1 1	0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 1 1 0	0 0 0 1 1 1 1 0 0 1 1 0 1 0 0	0 1 1 1 1 0 0 1	0 1 0 1	0 0 1 0 1	0 0 1 0 1 1 0 1 0 1 0 0 0 0 0
645/5	1 0 1 0 1 0 1	1 1	0 1 0 1 0 1	0 1 0 1 0	1 0 0	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0	0 0 0 1 1	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 1 0	0 0 0 0 1 1 1 0 0 1 0 1 1 0 0	0 1 1 1 0 0 0 1	0 0 0 1	0 0 1 0 0	0 0 0 0 1 1 0 1 0 1 0 0 0 0 0
645/6	0 0 1 0 1 0 1	0 0	0 1 0 1 0 1	0 1 0 1 0	1 0 0	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 0	0 1 1 1 0	0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1	0 0 0 0 1 1 0 0 0 1 0 1 1 0 0	0 1 1 1 1 0 0 1	1 0 0 1	0 0 1 1 0	0 0 0 0 1 1 0 0 0 0 0 0 0 1 0
645/14	0 0 0 0 1 1 1	1 1	0 1 0 1 1 1	0 0 1 1 1	1 0 0	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 1 0 0 0 0 1 0 0 0 1 1 1 0 0 1	0 0 0 1 0	0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0	0 0 0 0 1 1 1 0 0 1 0 1 1 0 0	0 0 0 0 0 0 0 0	0 0 1 1	0 1 0 1 0	0 0 0 0 0 1 0 0 0 0 0 0 0 1 0
688/1-07	0 1 0 0 0 1 0	0 0	0 0 0 0 0 0	0 0 1 1 1	0 1 0	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 1 1	0 1 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
688/4	0 1 0 0 0 1 0	0 0	0 0 0 0 1 0	0 0 0 0 1	0 1 1	0 0	0 0	0 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 0	? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
688/6	1 1 0 0 0 1 0	0 0	0 0 0 0 0 1	0 0 0 0 1	0 0 1	0 1	? ?	0 0 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 1 1	0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1	0 1 0 1	0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
688/7	1 1 0 0 0 1 0	0 0	0 0 0 0 0 0	0 0 0 0 1	0 0 1	0 1	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 1 0	? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
688/10	1 1 0 0 0 1 0	0 0	0 0 0 0 1 1	0 0 0 0 1	0 0 1	0 1	? ?	0 1 0 1 0
TRFK	0 0 0 0 0 0 1 0 0 0 0 1 0 0 1 1 0 0 0 0	0 0 0 1 1	0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 1	0 0 0 0 1 0 1 0 1 0 1 0 1 1 0	0 0 0 0 0 0 0 0	0 1 1 0	0 0 0 1 0	0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
688/11	1 1 0 0 0 1 0	0 0	0 1 0 0 1 0	0 1 0 1 1	0 1 0	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 1 1	0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 1 1 0	? ? ? ? ?	1 0 1 1 0 0 0 0 0 0 0 0 1 0 0
688/12	1 1 0 0 0 1 0	0 0	0 0 0 0 0 1	0 0 0 0 1	0 1 1	0 1	? ?	0 1 0 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 0 0 0	0 0 1 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 1 1 1	0 0 0 0 0 0 0 0	0 1 1 0	0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
688/13	1 0 0 0 0 1 0	1 1	0 1 0 0 1 1	0 1 0 1 1	1 1 1	0 0	0 0	0 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 1	? ? ? ? ?	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
688/15	0 0 1 0 0 1 0	1 0	? ? ? ? ?	0 0 0 0 1	0 1 1	0 0	? ?	? ? ? ?
TRFK	1 1 1 0 0 0 1 1 0 0 0 1 0 0 0 1 1 0 0 0	0 0 1 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 0 0 1 1 1 0 1 0 1 1 1 1 1 1	0 1 0 1 0 0 0 0	0 1 1 0	0 0 0 1 0	0 0 1 1 1 1 1 0 1 1 0 0 0 1 1
688/18	1 1 0 0 0 1 0	1 1	0 0 0 0 0 1	0 1 0 1 1	1 1 0	0 0	0 0	0 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 1 0 1 0 1 0 0 0 0 1 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1 0 0 1	0 1 1 0	0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
688/19	1 1 0 0 0 1 0	0 0	0 1 0 0 1 1	0 0 1 1 1	0 1 1	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 1 0 1	0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1 1 0 0	0 1 1 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
845/1	1 1 0 0 0 1 0	1 1	0 0 0 1 0 1	0 0 1 1 1	1 1 1	0 0	0 0	0 1 1 0 0

TRFK	0 0 0 0 0 0 1 0 0 0 0 1 0 0 1 1 1 0 0 1	0 0 0 1 1	0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0	0 0 0 1 1 0 1 1 0 1 1 0 1 1 1	0 0 1 0 1 1 0 1	0 1 1 0	0 0 0 0 1	0 0 0 0 1 1 0 0 0 0 0 0 0 1 0
845/2	0 1 0 0 0 1 0	0 0	0 1 1 1 0 1	0 0 1 1 1	0 1 1	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0	0 0 0 0 0	0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 1 0	0 1 0 0 1 0 1 1 0 1 1 0 1 0 0	0 0 1 0 1 0 0 0	0 1 1 0	0 0 0 0 1	0 0 0 0 0 1 0 0 0 0 0 0 0 0 1
845/3	1 1 0 0 0 1 0	1 0	0 1 1 1 1 0	0 0 1 1 1	1 1 1	0 0	0 0	0 1 1 0 0
TRFK	0 0 0 0 0 0 1 0 0 0 0 1 0 0 1 1 0 0 0 0	1 0 0 1 0	0 0 0 1 1 0 0 0 1 0 0 0 1 0 0 0 1 0	0 0 0 0 1 0 1 1 0 1 1 0 1 1 0	0 0 1 0 1 0 0 0	0 1 1 0	0 0 0 0 1	0 0 0 0 1 1 0 0 1 0 0 1 0 1 0
845/4	1 1 0 0 0 0 1	1 0	0 1 1 1 1 0	0 0 1 1 1	1 1 1	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 1 1	0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 1 0	? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
845/5	1 1 0 0 0 0 1	0 0	0 0 0 0 0 0	0 0 0 0 1	0 1 1	0 0	? ?	0 1 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0	0 0 0 0 0	0 0 0 0 0 0 0 0 1 0 0 0 1 1 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1 1 0 0	0 0 1 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
845/6	1 0 1 0 0 0 1	1 0	0 1 1 1 1 0	0 0 1 0 1	1 1 1	1 0	0 0	0 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0	0 0 0 0 0	0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 1 0	0 0 0 0 1 0 1 1 0 0 1 0 1 1 0	0 0 1 0 1 0 0 0	0 1 1 0	0 0 0 0 1	0 0 0 0 1 1 1 0 0 0 0 0 0 0 0
862/1	1 1 0 0 0 0 1	1 0	0 1 1 1 1 0	0 0 1 1 1	1 1 1	0 0	0 0	1 1 1 0 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 1 0	1 1 1 0 1 0 1 1 0 0 0 0 0 0 0 1 0 0	0 0 0 0 0 1 0 0 0 1 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 1 0	0 0 0 0 1	1 0 1 0 0 0 1 1 0 0 0 0 0 0 0
862/3	0 1 1 0 0 0 1	1 0	0 0 0 0 0 1	0 0 0 0 1	0 1 1	0 1	0 0	0 1 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0	0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
862/4	0 1 1 0 0 0 1	1 0	0 0 0 0 0 1	0 0 0 1 1	0 0 1	0 1	0 0	0 1 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1 1 0 0	0 0 1 0	0 0 1 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
862/5	1 0 1 0 0 0 1	1 0	0 1 0 1 1 0	0 0 1 1 1	1 1 1	0 0	0 0	1 0 1 0 0
TRFK	0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 1 1 0	0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 0 1 0	0 0 1 0 1 1 1 1 0 1 0 1 0 1 0	1 0 1 0 1 0 0 0	0 1 1 0	0 0 1 0 1	0 0 0 0 0 0 0 1 1 0 0 0 0 0 0
862/6	0 0 1 0 0 1 0	1 1	0 0 0 0 0 0	0 1 1 0 1	0 1 1	0 0	1 0	0 0 1 1 1
TRFK	0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1	0 0 1 1 0	0 0 0 0 0 0 0 0 1 0 0 1 1 0 1 0 1 0	0 0 0 1 1 1 1 1 0 1 0 1 0 0 0	0 0 1 0 1 0 0 0	0 1 1 0	0 0 0 0 1	0 0 0 0 1 0 0 1 1 1 0 0 0 0 0
862/7	0 0 1 0 0 1 0	1 1	0 0 0 0 0 0	0 1 0 1 1	1 1 1	0 0	0 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 1 0 0	0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 1 0 0	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
862/9	1 0 1 0 0 0 1	0 0	0 0 0 0 0 0	1 0 0 0 1	0 0 1	0 0	1 0	0 1 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1	0 0 1 1 0	0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 1 1 0	0 0 0 0 1 1 1 1 0 1 0 1 0 1 1	0 0 1 0 1 0 0 0	0 1 1 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
862/11	0 0 1 0 0 1 0	0 0	0 0 0 0 0 0	0 1 1 1 1	1 1 1	0 0	0 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 1 0	0 0 0 0 1 0 0 1 0 1 0 1 0 1 0	0 0 0 0 1 0 0 0	0 1 1 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
862/14	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 1 1	1 1 1	0 0	0 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 1	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
862/16	0 0 1 0 0 1 0	1 0	? ? ? ? ?	0 0 0 0 1	0 1 1	0 0	0 1	0 0 0 0 1
TRFK	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1	0 0 1 1 1	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 0 1	? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
862/20	0 0 1 0 0 1 0	1 1	0 0 0 0 0 0	0 0 0 0 1	0 1 1	0 0	? ?	0 0 1 0 1
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	? ? ? ? ? ? ? ? ?	0 0 0 1	0 0 0 0 0	? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
862/22	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 0 0 1	? ?	0 0	0 1	? ? ? ?

TRFK	0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0	0 0 0 0 0 0 0 1 1 0 0 1 0 1 0 0 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 1 0 0 1 0	0 0 1 1	0 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
570/1	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 0 1	0 1 1	0 0	0 0	0 0 0 1 1
TRFK	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0	0 0 1 1 0	1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0	0 0 1 0 0 1 0 0 0 1 0 1 0 0 1	0 0 0 0 0 0 0 0	0 1 0 0	0 0 0 0 0	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
570/2	1 0 1 0 0 0 1	0 0	0 0 0 0 0 0	0 1 0 0 1	0 0 1	0 0	1 0	? ? ? ?
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 1 0	0 0 0 0 0 0 0 1 1 0 0 1 0 1 0 0 1 1	0 0 0 0 1 1 1 0 1 1 0 0 1 1 0	0 1 0 1 0 0 0 0	0 0 1 1	0 1 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
597/1	0 0 1 0 0 1 0	0 0	0 0 0 0 0 0	0 0 1 0 1	1 1 1	0 0	0 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1	0 0 0 0 1 1 0 1 0 1 0 1 0 1 0	0 0 0 0 0 0 0 0	0 0 1 1	0 1 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
597/8	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 0 1	1 1 0	0 0	0 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1	0 0 1 0	0 0 1 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
597/12	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 0 1	0 1 1	0 0	0 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 1 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
597/15	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 0 0 1	0 0 1	0 0	1 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 1 1 0	0 0 0 0 0 0 0 0 1 0 1 1 1 0 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 1 0	? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
597/17	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 0 0 1	0 0 1	0 0	? ?	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	? ? ? ? ?	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 0 1 1 1 0 1 0 1 0 1 0 1 0	0 0 0 0 0 0 0 0	0 0 1 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
597/26	0 0 1 0 0 1 0	? ?	0 0 0 0 0 0	0 1 1 0 1	1 1 1	0 0	0 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 1 1 0	0 0 0 0 0 0 0 1 1 1 1 1 0 1 0 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 1 1 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
599/2	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 0 1	1 1 1	0 0	0 0	0 0 1 1 1
TRFK	? ?	0 0 0 0 0	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	0 0 1 1 1 1 0 0 0 1 0 1 0 0 0	0 0 0 0 0 0 0 0	0 1 1 0	0 0 0 0 1	0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
600/3	? ? ? ? ?	1 0	? ? ? ? ?	0 0 1 1 1	0 1 1	0 0	0 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	? ? ? ? ?	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
660/1	1 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 0 0 1	0 0 1	?	1 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	? ? ? ? ?	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
667/3	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 0 1	0 0 1	?	1 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 1 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 1 0 0 1 0 0	0 1 1 0	0 0 1 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
680/2	1 0 0 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 0 1	0 1 1	0 0	0 0	0 0 1 1 1
TRFK	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0	? ? ? ? ?	0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0	0 1 0 0 1 1 0 1 0 1 0 0 0 1 0	0 0 0 0 0 0 0 0	0 1 0 0	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
688/1-05	1 0 1 0 0 0 1	? ?	0 0 0 0 0 0	0 1 0 1 1	0 0 1	0 0	1 0	0 1 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1	0 0 0 0 0	0 0 0 0 0 0 1 1 0 1 1 0 1 0 1 1 1 1	0 0 0 0 1 0 1 0 0 1 0 1 0 0 0	0 0 0 0 0 1 0 0	0 0 1 0	0 0 0 0 1	1 1 0 0 0 0 0 0 0 0 0 0 0 0 1
691/1	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 1 0 1 1	0 1 1	0 0	0 0	0 1 0 1 0
TRFK	? ?	0 0 0 1 0	0 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 1 1	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	0 0 0 0 0 0 0 1	0 1 1 0	0 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
691/2	? ? ? ? ?	1 0	0 0 1 1 1 1	? ? ? ?	1 0 1	0 0	0 0	0 1 0 1 0
TRFK	0 0 1 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0	0 0 0 0 1	0 0 0 0 0 0 0 1 1 0 1 0 1 0 0 1 0 1	0 0 0 1 1 1 1 1 0 1 0 1 1 0 0	0 0 0 0 0 0 0 0	0 1 1 0	0 0 1 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
921/1	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 1 0 1 1	0 1 1	0 0	1 0	0 1 0 1 0

TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1	0 0 1 0 1 1 1 1 0 1 0 1 0 1 0	0 0 0 0 0 0 0 0	0 0 1 0	0 0 0 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
921/5	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 1 1	1 1 1	0 0	0 0	0 1 0 0 0
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0	0 0 0 1 0	0 0 1 0 0 1 0 0 1 1 0 0 1 0 0 0 0 0	0 1 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 1 0 0	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
TRFK 6/8.	1 0 1 0 0 0 1	0 0	0 0 0 0 0 1	1 0 0 0 1	0 0 1	0 0	1 0	0 1 0 1 0
TRFK	0 0 1 0 1 0 0 0 0 1 1 0 0 0 0 1 1 0 0 1	0 0 0 0 1	0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1	0 0 0 0 1 1 1 0 0 1 1 0 0 1 0	0 0 0 0 0 0 0 0	0 1 1 0	0 0 0 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
303/577	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 1 1 0 1	0 1 1	0 0	0 0	0 1 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 1 0	0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
301/4	1 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 0 1	0 1 1	0 0	0 1	0 1 0 1 0
TRFK K-	0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0	1 0 0 0 1	0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 1 0	0 0 0 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
purple	1 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 0 1	0 1 1	0 0	1 0	0 1 0 1 0
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 1 0 0 0	0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 1 0 0 0 1 0 0 0	0 0 0 0 0 0 0 0	0 1 1 0	0 0 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
GW Ejulu	0 0 1 0 0 0 0	0 0	0 0 0 0 0 0	0 0 1 1 0	0 1 1	0 0	0 0	0 1 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 0 1 0	0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 0 1 0	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
301/3	0 0 1 0 0 1 0	1 0	0 0 0 0 0 0	0 0 1 1 0	0 0 1	0 0	1 0	0 1 0 1 0
AHP	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	? ? ? ? ?	0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	? ? ? ? ?	? ? ? ? ?	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
S15/10	0 0 1 0 0 1 0	? ?	0 0 0 0 0 0	0 0 0 0 1	0 0 1	?	? ?	? ? ? ?
	0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 1 1 0	0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
BBK BB35	1 0 1 0 0 0 1	1 0	0 0 0 0 0 0	0 0 0 0 1	0 0 1	0 0	1 0	0 1 0 1 0
AHP	0 0 1 0 0 1 1 1 0 0 1 0 0 0 0 1 0 0 0 1	0 1 1 1 0	0 0 0 0 1 0 0 0 0 0 1 0 1 0 0 0 1 0	0 0 0 1 1 0 1 0 0 1 0 1 0 1 0	0 0 0 0 0 0 0 0	0 1 0 0	1 1 0 1 0	0 1 0 1 0 1 1 0 0 0 0 0 0 0 1
SC12/28	1 0 1 0 0 0 1	1 1	0 1 0 0 0 0	0 0 0 0 0	1 1 0	0 0	0 0	0 1 0 1 0
	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 1 0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0	0 0 0 0 1 1 1 0 1 1 0 1 0 1 0	0 0 0 0 0 0 0 0	0 1 0 0	0 0 1 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
TRFK 31/8	1 0 1 0 0 0 1	1 0	0 0 0 0 0 0	0 1 0 0 1	0 1 1	0 0	0 0	0 1 0 1 0
TRFK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1	0 1 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0	0 1 0 0	0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
301/2	1 0 1 0 0 0 1	1 0	0 0 0 0 0 0	0 0 0 0 1	0 0 1	0 0	1 0	0 1 0 1 0
EPK	0 0 1 0 0 1 0 1 0 0 1 0 0 0 1 0 1 0 0 1	0 1 0 1 1	0 1 0 0 1 0 0 1 0 1 0 1 1 0 0 0 1 0	0 0 1 0 1 0 1 0 0 1 0 1 0 1 0	0 0 1 0 1 0 0 0	0 1 0 0	0 0 0 1 1	1 1 0 1 1 0 0 0 0 0 0 0 0 0 1
TN14-3	1 0 0 0 0 1 0	1 1	0 1 1 0 0 0	0 1 0 0 0	1 1 1	0 0	0 0	0 1 0 0 0

Appendix V: Simulation Parameters Output from Cervus for Simulation of Mother

Input

Number of offspring:	10000
Number of candidate mothers:	12
Proportion of candidate mothers sampled:	0.16667
Proportion of loci typed:	0.5250000
Proportion of loci mistyped:	0.010000
Error rate in likelihood calculations:	0.05
Minimum number of typed loci:	1

Output

Confidence determined using:	LOD
Relaxed confidence level:	80%
Strict confidence level:	95%

Appendix VI: Confidence Level Analysis of Maternity

Mother alone:

Level	Confidence (%)	Critical LOD	Assignments	Assignment Rate
Strict	95.00	3.94	4	0%
Relaxed	80.00	3.00	35	0%
Unassigned			9965	100%
Total			10000	100%

Mother given known father:

Level	Confidence (%)	Critical LOD	Assignments	Assignment Rate
Strict	95.00	4.88	4	0%
Relaxed	80.00	3.25	50	1%
Unassigned			9950	99%

Appendix VII: Simulation Parameters Output from Cervus for Simulation of Father

Input

Number of offspring:	10000
Number of candidate mothers:	2
Proportion of candidate mothers sampled:	0.0833
Proportion of loci typed:	0.5250000
Proportion of loci mistyped:	0.010000
Error rate in likelihood calculations:	0.05
Minimum number of typed loci:	1

Output

Confidence determined using:		LOD
Relaxed confidence level:		80%
Strict confidence level:		95%
Total	*10000*	*100%*

Appendix VIII: Confidence Level Analysis of Paternity Assignment

Father alone:

Level	Confidence (%)	Critical LOD	Assignments	Assignment Rate
Strict	95.00	2.79	31	0%
Relaxed	80.00	0.58	658	7%
Unassigned			9342	93%
Total			10000	100%

Father given known mother:

Level	Confidence (%)	Critical LOD	Assignments	Assignment Rate
Strict	95.00	3.31	30	0%
Relaxed	80.00	0.22	819	8%
Unassigned			9181	92%
Total			10000	100%